PI 1129097

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

February 21, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/427,564*
FILING DATE: *November 20, 2002*
RELATED PCT APPLICATION NUMBER: *PCT/US03/37206*

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

T. WALLACE
Certifying Officer

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
REQUEST FOR FILING PROVISIONAL PATENT APPLICATION
Under 35 USC 111(b)
(Not for DESIGN cases)

Box: PROVISIONAL APPLICATION

Hon. Commissioner of Patents
Washington, D.C. 20231

PROVISIONAL APPLICATION
Under Rule 53(c)

Sir:

Herewith is a PROVISIONAL APPLICATION
Title: NOVEL GENE TARGETS AND LIGANDS THAT BIND THERETO FOR TREATMENT AND DIAGNOSIS OF COLON CARCINOMAS

Atty. Dkt.	PW 0291922	2002-30-0017P4
	M#	Client Ref

Date: November 20, 2002

Including:

1. Specification: 88 pages 2. ☐ Specification in non-English language 3. ☒ Drawings: 19 sheet(s)
4. The invention ☐ was ☒ was not made by, or under a contract with, an agency of the U.S. Government.
   If yes, Government agency/contact # = _____
5. ☐ Attached is an assignment and cover sheet. Please return the recorded assignment to the undersigned.
6. Small Entity Status ➡ ☒ is Not claimed ☐ is claimed (pre-filing confirmation **required**)
   NOTE: Do NOT File IDS!
7. ☐ Attached:

8. This application is made by the following named inventor(s) **(Double check instructions for accuracy.):**

	First	Middle Initial	Family Name
(1) Inventor	Karen		McLACHLAN
Residence	Solana Beach (City)	California (State/Foreign Country)	Great Britain (Country of Citizenship)
(2) Inventor	Dennis		GATELY
Residence	San Diego (City)	California (State/Foreign Country)	USA (Country of Citizenship)
(3) Inventor			
Residence	(City)	(State/Foreign Country)	(Country of Citizenship)
(4) Inventor			
Residence	(City)	(State/Foreign Country)	(Country of Citizenship)
(5) Inventor			
Residence	(City)	(State/Foreign Country)	(Country of Citizenship)

9. NOTE: FOR ADDITIONAL INVENTORS, check box ☐ and attach sheet (PAT102A) with same information regarding additional inventors.

	Large/Small Entity		Fee Code
10. Filing Fee	$160/$80	+160	114/214
11. If "non-English" box 2 is X'd, add Rule 17(k) processing fee	$130	+0	139
12. If "assignment" box 5 is X'd, add recording fee.	$40	+0	581
13.	TOTAL FEE ENCLOSED =	$160	

Our Deposit Account No. 03-3975
Our Order No. 037903 (C#) | 0291922 (M#)

00909

CHARGE STATEMENT. The Commissioner is hereby authorized to charge any fee specifically authorized hereafter, or any missing or insufficient fee(s) filed, or asserted to be filed, or which should have been filed herewith or concerning any paper filed hereafter, and which may be required under Rules 16-17 (missing or insufficient fee only) now or hereafter relative to this application or credit any overpayment, to our Account/Order Nos shown in the heading hereof for which purpose a duplicate copy of this sheet is attached.

Intellectual Property Group
By Atty: Robin L. Teskin Reg. No. 35,030
Sig: Fax:
Tel: 703-905-2200

Atty/Sec: RLT/af

NOTE: File in duplicate with 2 post card receipts (PAT-103) & attachments

NOVEL GENE TARGETS AND LIGANDS THAT BIND THERETO FOR TREATMENT AND DIAGNOSIS OF COLON CARCINOMAS

RELATED APPLICATIONS

This application relates to U.S. Provisional Serial No. 60/367,727 filed March 28, 2002 , U.S. Provisional Serial No. 60/381,328 filed May 20, 2002, and U.S. Provisional Serial No. 60/386, 747 filed June 10, 2002 each which are incorporated by reference in their entirety herein.

FIELD OF THE INVENTION

The present invention relates the identification of gene targets for treatment and diagnosis of cancers, especially colon or colorectal cancer, and other cancers wherein the subject genes are upregulated and the use thereof to express the corresponding antigen, and to produce ligands that specifically bind such antigen, e.g. monoclonal antibodies and small molecules.

DESCRIPTION OF RELATED ART

Colorectal cancers are among the most common cancers in men and women in the U.S. and are one of the leading causes of death. Other than surgical resection no other systemic or adjuvant therapy is available. Vogelstein and colleagues have described the sequence of genetic events that appear to be associated with the multistep process of colon cancer development in humans (Fearon and Vogelstein, 1990). An understanding of the molecular genetics of carcinogenesis, however, has not led to preventative or therapeutic measures. It can be expected that advances in molecular genetics will lead to better risk assessment and early diagnosis but colorectal cancers will remain a deadly disease for a majority of patients due to the lack of an adjuvant therapy. Adjuvant or systemic treatments are likely to arise from a better understanding of the autocrine factors responsible for the continued proliferation of cancer cells.

Endogenous gastrins and exogenous gastrins (other than tetragastrin) seem to promote the growth of established colon cancers in mice (Singh, et cl., 1986; Singh, et al., 1987; et al., 1984; Smith and Solomon, 1988; Singh, el cit., 1990; Rehfeld and van

Solinge, 1994) and promote carcinogen induced colon cancers in rats (Williamson et al., 1978; Karlin et al., 1985; Lamoste and Willems; 1988). Recent studies of Montag et al (1993) further support a possible co-carcinogenic role of gastrin in the initiation of tumors.

Many colon cancer cells express and secrete gastrin gene products (Dai et al., 1992; Kochinan et al., 1992; Finley et al., 1993; Van Solinge et al., 1993; Xu et al., 1994; Singh et al., 1994a; Hoosein et al., 1988; Hoosein et al., 1990) and bind gastrin-like peptides (Singh et al., 1986; Singh et al., 1987; Weinstock and Baldwin, 1988; Watson and Steele, 1994; Upp et al., 1989; Singh et al., 1985). In previous reports gastrin antibodies were either reported to inhibit (Hoosein et al., 1988; Hoosein et al, 1990) the growth of colon cancer cell lines in vitro.

However other investigators have had inconclusive results with colon cancer cell lines. A number of studies testing the effects of gastrin on cell proliferation of cancer cells have been performed (Sirinek et al., 1985; Kusyk et al., 1986; Watson et al., 1989). The results have varied widely. In one study, four different human cancer cell lines were tested for growth stimulation by pentagastrin and only one showed growth stimulation (Eggstein et al., 1991). Similarly in majority of the studies conducted to-date, mitogenic effects of gastrin have been demonstrated only on a very small percentage of colon cancer cell lines in vitro (Hoosein et al., 1988; Hoosein et al, 1990; Shrink et al, 1985; Kusyk et al, 1986; Guo et al, 1990; Ishizuka et al, 1994).

Since only a small percentage of established human colon cancer cell lines demonstrated a growth response to exogenous gastrins, investigators in this field came to believe that gastrin probably did not play a significant role in the growth of colon cancers. The recent discovery that human colon cancer cell lines and primary human colon cancers express the gastrin gene has sparked a renewed interest in a possible autocrine role of gastrin-like peptides in colon cancers. However, significant skepticism remains in the field, to date, regarding the importance of gastrin gene expression to the continued growth and tumorigenicity of colon cancers.

Thus, to-date, no systemic or adjuvant therapies have been developed for colon cancers, based on the knowledge that a significant percentage of human colon cancers express the gastrin gene. In fact, no adjuvant or systemic therapy has been developed for colon cancers that is based on the knowledge of the expression of other

growth factors such as TGF-alpha. or IGF-II, since none of the growth factors demonstrate a significant growth effect on majority of the colon cancer cell lines in culture.

At the present time the only systemic treatment available for colon cancer is chemotherapy. However, chemotherapy has not proven to be very effective for the treatment of colon cancers for several reasons, the most important of which is the fact that colon cancers express high levels of the MDR gene (that codes for multi-drug resistance gene products). The MDR gene products actively transport the toxic substances out of the cell before the chemotherapeutic agents can damage the DNA machinery of the cell. These toxic substances harm the normal cell populations more than they harm the colon cancer cells for the above reasons.

There is no effective systemic treatment for treating colon cancers other than surgically removing the cancers. In the case of several other cancers, including breast cancers, the knowledge of growth promoting factors (such as EGF, estradiol, IGF-II) that appear to be expressed or effect the growth of the cancer cells, has been translated for treatment purposes. But in the case of colon cancers this knowledge has not been applied and therefore the treatment outcome for colon cancers remains bleak.

Antisense RNA technology has been developed as an approach to inhibiting gene expression, particularly oncogene expression. An "antisense" RNA molecule is one which contains the complement of, and can therefore hybridize with, protein-encoding RNAs of the cell. It is believed that the hybridization of antisense RNA to its cellular RNA complement can prevent expression of the cellular RNA, perhaps by limiting its translatability. While various studies have involved the processing of RNA or direct introduction of antisense RNA oligonucleotides to cells for the inhibition of gene expression (Brown, et al., 1989; Wickstrom, et al., 1988; Smith, et al., 1986; Buvoli, et al., 1987), the more common means of cellular introduction of antisense RNAs has been through the construction of recombinant vectors which will express antisense RNA once the vector is introduced into the cell.

A principle application of antisense RNA technology has been in connection with attempts to affect the expression of specific genes. For example, Delauney, et al. have reported the use antisense transcripts to inhibit gene expression in transgenic plants (Delauney, et al., 1988). These authors report the down-regulation of

chloramphenicol acetyl transferase activity in tobacco plants transformed with CAT sequences through the application of antisense technology.

Antisense technology has also been applied in attempts to inhibit the expression of various oncogenes. For example, Kasid, et al., 1989, report the preparation of recombinant vector construct employing Craf-1 cDNA fragments in an antisense orientation, brought under the control of an adenovirus 2 late promoter. These authors report that the introduction of this recombinant construct into a human squamous carcinoma resulted in a greatly reduced tumorigenic potential relative to cells transfected faith control sense transfectants. Similarly, Prochownik, et al., 1988, have reported the use of Cmiyc antisense constructs to accelerate differentiation and inhibit G.sub.1 progression in Friend Murine Erythroleukemia cells. In contrast, Khokha, et al., 1989, discloses the use of antisense RNAs to confer oncogenicity on 3T3 cells, through the use of antisense RNA to-reduce murine tissue inhibitor or metalloproteinases levels.

Antisense methodology takes advantage of the fact that nucleic acids tend to pair with "complementary" sequences. By complementary, it is meant that polynucleotides are those which are capable of base-pairing according to the standard Watson-Crick complementary rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

Targeting double-stranded (ds) DNA with polynucleotides leads to triple-helix formation; targeting RNA will lead to double-helix formation. Antisense polynucleotides, when introduced into a target cell, specifically bind to their target polynucleotide and interfere with transcription, RNA processing, transport, translation and/or stability. Antisense RNA constructs, or DNA encoding such antisense RNA's, may be employed to inhibit gene transcription or translation or both within a host cell, either in vitro or in vivo, such as within a host animal, including a human subject.

Throughout this application, the term "expression vector or construct" is meant to include any type of genetic construct containing a nucleic acid coding for a gene

product in which part or all of the nucleic acid encoding sequence is capable of being transcribed. The transcript may be translated into a protein but it need not be. Thus, in certain embodiments, expression includes both transcription of a gene and translation of mRNA into a gene product. In other embodiments, expression only includes transcription of the nucleic acid encoding a gene of interest.

The nucleic acid encoding a gene product is under transcriptional control of a promoter. A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a gene. The phrase "under transcriptional control" means that the promoter is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the gene.

The term promoter is used to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II. Much of the thinking about how promoters are organized derives from analyses of several viral promoters, including those for the HSV thymidine kinase (tk) and SV40 early transcription units. These studies, augmented by more recent work, have shown that promoters are composed of discrete functional modules, each consisting of approximately 7-20 bp of DNA, and containing one or more recognition sites for transcriptional activator or repressor proteins.

At least one module in each promoter functions to position the start site for RNA synthesis. The best known example of this is the TATA box, but in some promoters lacking a TATA box, such as the promoter for the mammalian terminal deoxynucleotidyl transferase gene and the promoter for the SV40 late genes, a discrete element overlying the start site itself helps to fix the place of initiation.

Additional promoter elements regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the tk promoter, the spacing between promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter, it appears that individual elements can function either

cooperatively or independently to activate transcription.

The particular promoter that is employed to control the expression of a nucleic acid encoding a particular gene is not believed to be important, so long as it is capable of expressing the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell. Generally speaking, such a promoter might include either a human or viral promoter.

In various instances, the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter and the Rous sarcoma virus long terminal repeat can be used to obtain high-level expression of the gene of interest. The use of other viral or mammalian cellular or bacterial phage promoters which are well known in the art to achieve expression of a gene of interest is contemplated as well, provided that the levels of expression are sufficient for a given purpose.

By employing a promoter with well-known properties, the level and pattern of expression of the gene product following transfection can be optimized. Further, selection of a promoter that is regulated in response to specific physiologic signals can permit inducible expression of the gene product. Several elements/promoters which may be employed, in the context of the present invention, to regulate the expression of the gene of interest are listed below. This list is not intended to be exhaustive of all the possible elements involved in the promotion of gene expression but, merely, to be exemplary thereof.

Enhancers were originally detected as genetic elements that increased transcription from a promoter located at a distant position on the same molecule of DNA. This ability to act over a large distance had little precedent in classic studies of prokaryotic transcriptional regulation. Subsequent work showed that regions of DNA with enhancer activity are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins.

The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a

particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

Viral promoters, cellular promoters/enhancers and inducible promoters/enhancers that could be used in combination with the nucleic acid encoding a gene of interest in an expression construct. Some examples of enhancers include Immunoglobulin Heavy Chain; Immunoglobulin Light Chain; T-Cell Receptor; HLA DQ a and DQ b b-Interferon; Interleukin-2; Interleukin-2 Receptor: Gibbon Ape Leukemia Virus; MHC Class II 5 or HLA-DRa; b-Actin; Muscle Creatine Kinase; Prealbumin (Transthyretin); Elastase I; Metallothionein; Collagenase, Albumin Gene; a-Fetoprotein; a-Globin; b-Globin; c-fos: c-HA-ras; Insulin Neural Cell Adhesion Molecule (NCAM); al-Antitrypsin; H2B (TH2B) Histone; Mouse or Type I Collagen; Glucose-Regulated Proteins (GRP94 and GRP78); Rat Growth Hormone; Human Serum Amyloid A (SAA); Troponin I (TN I); Platelet-Derived Growth Factor; Duchenne Muscular Dystrophy; SV40 or CMV; Polyoma; Retroviruses; Papilloma Virus; Hepatitis B Virus; Human Immunodeficiency Virus. Inducers such as phorbol ester (TFA) heavy metals; glucocorticoids; poly (rl)X; poly(rc); Ela; H.sub.2 O.sub.2; IL 1; Interferon, Newcastle Disease Virus; A23187; IL-6; Serum; SV40 Large T Antigen; FMA; thyroid Hormone; could be used. Additionally, any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of the gene. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

In certain instances, the expression construct will comprise a virus or engineered construct derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal et al., 1986: Temin, 1986). The first viruses used as gene vectors were DNA viruses including the papoviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal et al., 1986) and adenoviruses (Ridgeway, 1988; Baichwal et al., 1986). These have a relatively low

capacity for foreign DNA sequences and have a restricted host spectrum. Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kB of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed. Also often another element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

It is understood in the art that to bring a coding sequence under the control of a promoter, or operatively linking a sequence to a promoter, one positions the 5' end of the transcription initiation site of the transcriptional reading frame of the protein between about 1and about 50 nucleotides "downstream" of (i.e., 3' of) the chosen promoter. In addition, where eukaryotic expression is contemplated, one will also typically desire to incorporate into the transcriptional unit which includes the cotransporter protein, an appropriate polyadenylation site (e.g., 5'-AATAAA-3') if one was not contained within the original cloned segment. Typically, the poly A addition site is placed about 30 to 2000 nucleotides "downstream" of the termination site of the protein at a position prior to transcription termination.

The above background references are part of the present invention insofar as they are applicable to the invention described herein. Hence there are no effective and specific ways of treating or diminishing the growth of colorectal cancer to date

Therefor, there exists a significant need for the identification of novel gene targets for the treatment and diagnosis of colon or colorectal cancer, especially given the huge human toll caused by this disease annually.

OBJECTS OF THE INVENTION

It is an object of the invention to identify novel gene targets for treatment and the diagnosis of cancer, especially colon or colorectal cancer.

It is a specific object of the invention to develop novel therapies for treatment of cancer, preferably colon cancer involving the administration of anti-sense oligonucleotides corresponding to gene targets that are expressed by certain colon or colorectal cancers.

It is another specific object of the invention to provide the antigens expressed by genes that are expressed by malignant tissues, e.g., colon or colorectal cancers.

It is another specific object of the invention to produce ligands that bind antigens expressed by certain cancers, especially colon or colorectal cancers, especially monoclonal antibodies.

It is another specific object of the invention to provide novel therapeutic regimens for the treatment of cancer, preferably colon cancer that involve the administration of antigens expressed by certain colon or colorectal cancers, alone or in combination with adjuvants that elicit an antigen-specific cytotoxic T-cell lymphocyte response against cancer cells that express such antigen.

It is another object of the invention to provide novel therapeutic regimens for the treatment of cancer, preferably colon or colorectal cancer that involve the administration of ligands, especially monoclonal antibodies that specifically bind novel antigens that are expressed by certain cancer tissues including colon cancer tissues.

It is another object of the invention to provide a novel method for diagnosis of cancer, preferably colon or colorectal cancer, by using ligands, e.g., monoclonal antibodies, that specifically bind to antigens that are expressed by cancers including certain colon or colorectal cancers, in order to detect whether a subject has or is at increased risk of developing colon or colorectal cancer.

It is another object of the invention to provide a novel method of detecting persons having, or at increased risk of developing certain types of cancers, including colon cancer by use of labeled DNAs that hybridize to novel gene targets expressed by certain cancers, including colon cancers.

It is yet another object of the invention to provide diagnostic test kits for the detection of persons having or at increased risk of developing certain cancer, including colon cancer that comprise a ligand, e.g., monoclonal antibody that specifically binds to an antigen expressed by certain colon cancers, and a detectable label, e.g., a radiolabel or fluorophore.

It is another object of the invention to provide diagnostic kits for detection of persons having or at risk of developing certain cancers, including colon cancer that comprise DNA primers or probes specific for novel gene targets expressed by colon cancers, and a detectable label, e.g. radiolabel or fluorophore.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 contains the gene expression profile determined using the Gene Logic datasuite for a DNA sequence overexpressed in colon tumor tissue (Genbank Accession W91975).

Figure 2 contains the gene expression profile determined using the Gene Logic datasuite for a DNA sequence overexpressed in colon tumor tissue (Genbank Accession A1694242).

Figure 3 contains the gene expression profile determined using the Gene Logic datasuite for a DNA sequence overexpressed in colon tumor tissue (Genbank Accession A7620111).

Figure 4 contains the gene expression profile determined using the Gene Logic datasuite for a DNA sequence overexpressed in colon tumor tissue (Genbank Accession AA813827).

Figure 5 contains expression data for three genes, CICO1, CICO2 and CICO3 identified as being overexpressed in colon cancer tissue.

Figure 6-11 contains E-Northern expression data for genes identified as being overexpressed in colon cancer.

Figure 12-16 contains the results of PCR expression analysis which analyzed the expression of CHEM1 in various panels of tissues.

Figure 17 contains Enorthern expression results for the NM_021246 gene.

Figure 18 contains Enorthern expression results for the A7821606 gene.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to the identification of genes which are to be specifically expressed and upregulated in certain cancers, including colon or colorectal tumors. This was determined using the Gene Logic datasuite or Celera database and by screening malignant colon tumor tissues as described in detail infra.

In particular, the present invention involves the discovery that certain genes, the nucleic acid sequences and predicted coding sequences of which are identified infra are specifically expressed in certain malignant tissues including colon or colorectal tumor tissues.

The particular genes and therapy nucleic acid and corresponding protein coding sequences which are the subject of this invention are disclosed in the examples.

Such therapies will involve the synthesis of oligonucleotides having sequences in the antisense orientation relative to the genes identified by the present inventors which are specifically expressed by malignant tissues, including colon or colorectal tumors. Suitable therapeutic antisense oligonucleotides will typically vary in length from two to several hundred nucleotides in length, more typically about 50-70 nucleotides in length. These antisense oligonucleotides may be administered as naked DNAs or in protected forms, e.g., encapsulated in liposomes. The use of liposomal or other protected forms may be advantageous as it may enhance *in vivo* stability and delivery to target sites, i.e., colon tumor cells.

Also, the subject novel genes may be used to design novel ribozymes that target the cleavage of the corresponding mRNAs in colon and other tumor cells. Similarly, these ribozymes may be administered in free (naked) form or by the use of delivery systems that enhance stability and/or targeting, e.g., liposomes. Ribozymal and antisense therapies used to target genes that are selectively expressed by cancer cells are well known in the art.

Also, the present invention embraces the administration of use of DNAs that hybridize to the novel gene targets identified *infra*, attached to therapeutic effector moieties, e.g., radiolabels, e.g., yttrium, iodine, cytotoxins, cytotoxic enzymes, in order to selectively target and kill cells that express these genes, i.e., colon tumor cells.

Still further, the present invention encompasses non-nucleic acid based therapies. Particularly, the invention encompasses the use of the nucleic acid sequences disclosed in the examples, for the expression of the corresponding antigens. It is anticipated that these antigens may be used as therapeutic or prophylactic anti-tumor vaccines. For example, a particular contemplated application of these antigens involves their administration with adjuvants that induce a cytotoxic T lymphocyte response. An especially preferred adjuvant developed by the Assignee of this application, IDEC Pharmaceuticals Corporation, is disclosed in U.S. Patent Nos. 5,709,860, 5,695,770, and 5,585,103, the disclosures of which are incorporated by reference in their entirety. In particular, the use of this adjuvant to promote CTL responses against prostate and papillomavirus related human colon cancer has been suggested.

Also, administration of the subject novel antigens in combination with an adjuvant may result in a humoral immune response against such antigens, thereby delaying or preventing the development of cancers associated with the overexpression thereof, e.g., colon cancer.

Essentially, these embodiments of the invention will comprise administration of one or both of the subject novel colon cancer antigens, ideally in combination with an adjuvant, e.g., PROVAX®, which comprises a microfluidized adjuvant containing Squalene, Tween and Pluronic, in an amount sufficient to be therapeutically or prophylactically effective. A typical dosage will range from 50 to 20,000 mg/kg body weight, have typically 100 to 5000 mg/kg body weight.

Alternatively, the subject tumor-associated antigens may be administered with other adjuvants, e.g., ISCOMS, DETOX, SAF, Freund's adjuvant, Alum, Saponin, among others.

Yet another embodiment of the invention will comprise the preparation of monoclonal antibodies against the antigens encoded by the DNA sequences disclosed in the examples which are expressed specifically by certain malignant tissues including colon or colorectal tumor tissues. Such monoclonal antibodies will be produced by conventional methods and include human monoclonal antibodies, humanized monoclonal antibodies, chimeric monoclonal antibodies, single chain antibodies, e.g., scFv's and antigen-binding antibody fragments such as Fabs, 2 Fabs, and Fab' fragments. Methods for the preparation of monoclonal antibodies and fragments thereof, e.g., by pepsin or papain-mediated cleavage are well known in the art. In general, this will comprise immunization of an appropriate (non-homologous) host with the subject colon cancer antigens, isolation of immune cells therefrom, use of such immune cells to make hybridomas, and screening for monoclonal antibodies that specifically bind to either of such antigens.

These monoclonal antibodies and fragments will be useful for passive anti-tumor immunotherapy, or may be attached to therapeutic effector moieties, e.g., radiolabels, cytotoxins, therapeutic enzymes, agents that induce apoptosis, in order to provide for targeted cytotoxicity, i.e., killing of human colon tumor cells. Given the fact that the subject genes are apparently not significantly expressed by many normal tissues this should not result in significant adverse side effects (toxicity to non-target tissues).

In this embodiment, such antibodies or fragments will be administered in labeled or unlabeled form, alone or in combination with other therapeutics, e.g., chemotherapeutics such

as progestin, EGFR, Taxol, etc. The administered composition will include a pharmaceutically acceptable carrier, and optionally adjuvants, stabilizers, etc., used in antibody compositions for therapeutic use.

Preferably, such monoclonal antibodies will bind the target antigens with high affinity, e.g., possess a binding affinity (Kd) on the order of 10^{-6} to 10^{-10} M.

As noted, the present invention also embraces diagnostic applications that provide for detection of the colon or colorectal tumor specific genes disclosed herein. Essentially, this will comprise detecting the expression of one or more of these genes at the DNA level or at the protein level.

At the DNA level, expression of the subject genes will be detected by known DNA detection methods, e.g., Northern blot hybridization, strand displacement amplification (SDA), catalytic hybridization amplification (CHA), and other known DNA detection methods. Preferably, a cDNA library will be made from colon cells obtained from a subject to be tested for colon cancer by PCR using primers corresponding to either or both of the novel genes disclosed in this application.

The presence or absence of cancer associated with the genes disclosed infra will be determined based on whether PCR products are obtained, and the level of expression. The levels of expression of such PCR product may be quantified in order to determine the prognosis of a particular colon cancer patient (as the levels of expression of the PCR product likely will increase as the disease progresses.) This may provide a method of monitoring the status of a cancer patient, e.g., colon cancer patient. Of course, suitable controls will be effected.

Alternatively, the status of a subject to be tested for colon or other cancer associated by overexpression of a gene disclosed infra may be evaluated by testing biological fluids, e.g., blood, urine, colon tissue, with an antibody or antibodies or fragment that specifically binds to the novel colon tumor antigens disclosed herein.

Methods of using antibodies to detect antigen expression are well known and include ELISA, competitive binding assays, etc. In general, such assays use an antibody or antibody fragment that specifically binds the target antigen directly or indirectly bound to a label that provides for detection, e.g., a radiolabel enzyme, fluorophore, etc.

For examples, patients which test positive for the presence of the antigen on colon cells will be diagnosed as having or being at increased risk of developing colon cancer.

Additionally, the levels of antigen expression may be useful in determining patient status, i.e., how far the disease has advanced (stage of particular cancer associated with overexpression of the particular gene).

As noted, the present invention provides novel genes and corresponding antigens that correlate to human colon cancer. The present invention also embraces variants thereof. By "variants" is intended sequences that are at least 75% identical thereto, more preferably at least 85% identical, and most preferably at least 90% identical when these DNA sequences are aligned to the subject DNAs or a fragment thereof having a size of at least 50 nucleotides. This includes in particular allelic variants of the subject genes.

Also, the present invention provides for primer pairs that result in the amplification DNAs encoding the subject novel genes or a portion thereof in an mRNA library obtained from a desired cell source, typically human colon cell or tissue sample. Typically, such primers will be on the order of 12 to 50 nucleotides in length, and will be constructed such that they provide for amplification of the entire or most of the target gene.

Also, the invention embraces the antigens encoded by the subject DNAs or fragments thereof that bind to or elicit antibodies specific to the full-length antigens. Typically, such fragments will be at least 10 amino acids in length, more typically at least 25 amino acids in length.

As noted, the subject genes are expressed in a majority of colon tumor samples tested. Additionally, some of these genes are upregulated in other cancers. The invention further contemplates the identification of other cancers that express such genes and the use thereof to detect and treat such cancers. For example, the subject genes or variants thereof may be expressed on other cancers, e.g., breast, pancreas, lung or colon cancers. Essentially, the present invention embraces the detection of any cancer wherein the expression of the subject novel genes or variants thereof correlate to a cancer or an increased likelihood of cancer.

"Isolated" refers to any human protein that is not in its normal cellular millieu. This includes by way of example compositions comprising recombinant protein, pharmaceutical compositions comprising purified protein, diagnostic compositions comprising purified protein, and isolated protein compositions comprising protein. In preferred embodiments, an isolated protein will comprise a substantially pure protein, in that it is substantially free of other proteins, preferably that is at least 90% pure, that comprises the amino acid sequence contained in the figures herein or natural homologues or mutants having essentially the same

sequence. A naturally occurring mutant might be found, for instance, in tumor cells expressing a gene encoding a mutated protein sequence.

"Native human protein" refers to a protein that comprises the amino acid sequence of the protein expressed in its endogenous environment, i.e., a human colon or colorectal tumor tissue.

"Native non-human primate protein" refers to a protein that is a non-human primate homologue of the protein having the amino acid sequence discussed in the examples. Given the phylogenetic closeness of humans to other primates, it is anticipated that human and non-human proteins expressed by the genes disclosed in the examples will have non-human primate counterparts that possess amino acid sequences that are highly similar, probably on the order of 95% sequence identity or higher.

"Isolated human or non-human primate nucleic acid molecule or sequence" refers to a nucleic acid molecule that encodes human protein which is not in its normal human cellular millieu, e.g., is not comprised in the human or non-human primate chromosomal DNA. This includes by way of example vectors that comprise a nucleic acid molecule, a probe that comprises a gene nucleic acid sequence directly or indirectly attached to a detectable moiety, e.g. a fluorescent or radioactive label, or a DNA fusion that comprises a nucleic acid molecule encoding a colon antigen according to the invention fused at its 5' or 3' end to a different DNA, e.g. a promoter or a DNA encoding a detectable marker or effector moiety. A preferred nucleic acid sequence encodes a human protein having the nucleic acid sequence in disclosed in the examples. Also included are natural homologues or mutants having substantially the same sequence. Naturally occurring homologies that are degenerate would encode the same protein as discussed infra in the examples, but would include nucleotide differences that do not change the corresponding amino acid sequence. Naturally occurring mutants might be found in tumor cells, wherein such nucleotide differences result in a mutant protein. Naturally occurring homologues containing conservative substitutions are also encompassed.

"Variant of human or non-human primate protein" refers to a protein possessing an amino acid sequence that possess at least 90% sequence identity, more preferably at least 91% sequence identity, even more preferably at least 92% sequence identity, still more preferably at least 93% sequence identity, still more preferably at least 94% sequence identity, even more preferably at least 95% sequence identity, still more preferably at least 96%

sequence identity, even more preferably at least 97% sequence identity, still more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity, to the corresponding native human or non-human primate protein wherein sequence identity is as defined infra. Preferably, this variant will possess at least one biological property in common with the human or non-human protein.

"Variant of human or non-human primate nucleic acid molecule or sequence" refers to a nucleic acid sequence that possesses at least 90% sequence identity, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, still more preferably at least 94%, even more preferably at least 95%, still more preferably at least 96%, even more preferably at least 97%, even more preferably at least 98% sequence identity, and most preferably at least 99% sequence identity, to the corresponding native human or non-human primate nucleic acid sequence, wherein "sequence identity" is as defined infra.

"Fragment of human or non-human primate nucleic acid molecule or sequence" refers to a nucleic acid sequence corresponding to a portion of the native human nucleic acid sequence discussed infra in the examples or a primate native non-human homolog molecule, wherein said portion is at least about 50 nucleotides in length, or 100, more preferably at least 200 or 300 nucleotides in length.

"Antigenic fragments of colon or colorectal" refer to polypeptides corresponding to a fragment of colon antigen encoded by any of the genes disclosed infra or a variant or homologue thereof that when used itself or attached to an immunogenic carrier that elicits antibodies that specifically bind the protein. Typically such antigenic fragments will be at least 20 amino acids in length.

Sequence identity or percent identity is intended to mean the percentage of the same residues shared between two sequences, referenced to the human DNA or amino acid sequences disclosed infra, when the two sequences are aligned using the Clustal method [Higgins et al, Cabios 8:189-191 (1992)] of multiple sequence alignment in the Lasergene biocomputing software (DNASTAR, INC, Madison, WI). In this method, multiple alignments are carried out in a progressive manner, in which larger and larger alignment groups are assembled using similarity scores calculated from a series of pairwise alignments. Optimal sequence alignments are obtained by finding the maximum alignment score, which is the average of all scores between the separate residues in the alignment, determined from a residue weight table representing the probability of a given amino acid change occurring in

two related proteins over a given evolutionary interval. Penalties for opening and lengthening gaps in the alignment contribute to the score. The default parameters used with this program are as follows: gap penalty for multiple alignment=10; gap length penalty for multiple alignment=10; k-tuple value in pairwise alignment=1; gap penalty in pairwise alignment=3; window value in pairwise alignment=5; diagonals saved in pairwise alignment=5. The residue weight table used for the alignment program is PAM250 [Dayhoffet al., in Atlas of Protein Sequence and Structure, Dayhoff, Ed., NDRF, Washington, Vol. 5, suppl. 3, p. 345, (1978)].

Percent conservation is calculated from the above alignment by adding the percentage of identical residues to the percentage of positions at which the two residues represent a conservative substitution (defined as having a log odds value of greater than or equal to 0.3 in the PAM250 residue weight table). Conservation is referenced to a human gene of the invention when determining percent conservation with a non-human gene and when determining percent conservation. Conservative amino acid changes satisfying this requirement are: R-K; E-D, Y-F, L-M; V-I, Q-H.

Polypeptide Fragments

The invention provides polypeptide fragments of the disclosed proteins. Polypeptide fragments of the invention can comprise at least 8, more preferably at least 25, still more preferably at least 50 amino acid residues of human or non-human primate gene according to the invention or an analogue thereof. More particularly such fragment will comprise at least 75, 100, 125, 150, 175, 200, 225, 250, 275 residues of the polypeptide encoded by gene the subject genes which are specifically expressed by certain human colon or colorectal as well as some other tumor tissues. Even more preferably, the protein fragment will comprise the majority of the native protein colon or colorectal protein, i.e. at least about 100 contiguous residues of the native colon or colorectal protein antigen.

Biologically Active Variants

The invention also encompasses biologically active mutants of protein colon or colorectal proteins according to the invention, which comprise an amino acid sequence that is at least 80%, more preferably 90%, still more preferably 95-99% similar to the subject tumor-associated, e.g., colon cancer associated proteins.

Guidance in determining which amino acid residues can be substituted, inserted, or deleted without abolishing biological or immunological activity can be found using computer programs well known in the art, such as DNASTAR software. Preferably, amino acid changes in protein variants are conservative amino acid changes, *i.e.*, substitutions of similarly charged or uncharged amino acids. A conservative amino acid change involves substitution of one of a family of amino acids which are related in their side chains. Naturally occurring amino acids are generally divided into four families: acidic (aspartate, glutamate), basic (lysine, arginine, histidine), non-polar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine) amino acids. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids.

A subset of mutants, called muteins, is a group of polypeptides in which neutral amino acids, such as serines, are substituted for cysteine residues which do not participate in disulfide bonds. These mutants may be stable over a broader temperature range than native secreted proteins. See Mark *et al.*, U.S. Patent 4,959,314.

It is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the biological properties of the resulting secreted protein or polypeptide variant.

Human or non-human primate protein variants include glycosylated forms, aggregative conjugates with other molecules, and covalent conjugates with unrelated chemical moieties. Also, protein variants also include allelic variants, species variants, and muteins. Truncations or deletions of regions which do not affect the differential expression of the protein gene are also variants. Covalent variants can be prepared by linking functionalities to groups which are found in the amino acid chain or at the N- or C-terminal residue, as is known in the art.

It will be recognized in the art that some amino acid sequence of the proteins of the invention can be varied without significant effect on the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there are critical areas on the protein which determine activity. In general, it is possible to replace residues that form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant if the

alteration occurs at a non-critical region of the protein. The replacement of amino acids can also change the selectivity of binding to cell surface receptors. Ostade et al., *Nature* 361:266-268 (1993) describes certain mutations resulting in selective binding of TNF-alpha to only one of the two known types of TNF receptors. Thus, the polypeptides of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation.

The invention further includes variations of the protein subject colon or colorectal which show comparable expression patterns or which include antigenic regions. Such protein mutants include deletions, insertions, inversions, repeats, and type substitutions. Guidance concerning which amino acid changes are likely to be phenotypically silent can be found in Bowie, J.U., et al., "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990).

Of particular interest are substitutions of charged amino acids with another charged amino acid and with neutral or negatively charged amino acids. The latter results in proteins with reduced positive charge to improve the characteristics of the disclosed protein. The prevention of aggregation is highly desirable. Aggregation of proteins not only results in a loss of activity but can also be problematic when preparing pharmaceutical formulations, because they can be immunogenic. (Pinckard et al., *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins et al., *Diabetes* 36:838-845 (1987); Cleland et al., *Crit. Rev. Therapeutic Drug Carrier Systems* 10:307-377 (1993)).

Amino acids in the polypeptides of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244: 1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as binding to a natural or synthetic binding partner. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J Mol. Biol.* 224:899-904 (1992) and de Vos et al. *Science* 255: 306-312 (1992)).

As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein. Of course, the number of amino acid substitutions a skilled artisan would make depends on many

factors, including those described above. Generally speaking, the number of substitutions for any given polypeptide will not be more than 50, 40, 30, 25, 20, 15, 10, 5 or 3.

Fusion Proteins

Fusion proteins comprising proteins or polypeptide fragments of the subject colon or colorectal proteins can also be constructed. Fusion proteins are useful for generating antibodies against amino acid sequences and for use in various assay systems. For example, fusion proteins can be used to identify proteins which interact with a protein of the invention or which interfere with its biological function. Physical methods, such as protein affinity chromatography, or library-based assays for protein-protein interactions, such as the yeast two-hybrid or phage display systems, can also be used for this purpose. Such methods are well known in the art and can also be used as drug screens. Fusion proteins comprising a signal sequence and/or a transmembrane domain of a protein according to the invention or a fragment thereof can be used to target other protein domains to cellular locations in which the domains are not normally found, such as bound to a cellular membrane or secreted extracellularly.

A fusion protein comprises two protein segments fused together by means of a peptide bond. Amino acid sequences for use in fusion proteins of the invention can utilize any of the amino acid sequences or encoded by the nucleotide sequences disclosed infra, or can be prepared from biologically active variants or fragment of said protein sequence, such as those described above. The first protein segment can consist of a full-length protein or a variant or fragment thereof. As noted, these fragments may range in size from about 8 amino acids up to the full length of the protein.

The second protein segment can be a full-length protein or a polypeptide fragment. Proteins commonly used in fusion protein construction include ß-galactosidase, ß-glucuronidase, green fluorescent protein (GFP), autofluorescent proteins, including blue fluorescent protein (BFP), glutathione-S-transferase (GST), luciferase, horseradish peroxidase (HRP), and chloramphenicol acetyltransferase (CAT). Additionally, epitope tags can be used in fusion protein constructions, including histidine (His) tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Other fusion constructions can include maltose binding protein (MBP), S-tag, Lex a DNA binding domain

(DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP 16 protein fusions.

These fusions can be made, for example, by covalently linking two protein segments or by standard procedures in the art of molecular biology. Recombinant DNA methods can be used to prepare fusion proteins, for example, by making a DNA construct which comprises a coding sequence encoding an amino acid sequence according to the invention in proper reading frame with a nucleotide encoding the second protein segment and expressing the DNA construct in a host cell, as is known in the art. Many kits for constructing fusion proteins are available from companies that supply research labs with tools for experiments, including, for example, Promega Corporation (Madison, WI), Stratagene (La Jolla, CA), Clontech (Mountain View, CA), Santa Cruz Biotechnology (Santa Cruz, CA), MBL International Corporation (MIC; Watertown, MA), and Quantum Biotechnologies (Montreal, Canada; 1-888-DNA-KITS).

Proteins, fusion proteins, or polypeptides of the invention can be produced by recombinant DNA methods. For production of recombinant proteins, fusion proteins, or polypeptides, a sequence listing encoding one of the subject colon or colorectal proteins can be expressed in prokaryotic or eukaryotic host cells using expression systems known in the art. These expression systems include bacterial, yeast, insect, and mammalian cells.

The resulting expressed protein can then be purified from the culture medium or from extracts of the cultured cells using purification procedures known in the art. For example, for proteins fully secreted into the culture medium, cell-free medium can be diluted with sodium acetate and contacted with a cation exchange resin, followed by hydrophobic interaction chromatography. Using this method, the desired protein or polypeptide is typically greater than 95% pure. Further purification can be undertaken, using, for example, any of the techniques listed above.

It may be necessary to modify a protein produced in yeast or bacteria, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain a functional protein. Such covalent attachments can be made using known chemical or enzymatic methods.

Human or non-human primate proteins according to the invention or polypeptide of the invention can also be expressed in cultured host cells in a form which will facilitate purification. For example, a protein or polypeptide can be expressed as a fusion protein

comprising, for example, maltose binding protein, glutathione-S-transferase, or thioredoxin, and purified using a commercially available kit. Kits for expression and purification of such fusion proteins are available from companies such as New England BioLabs, Pharmacia, and Invitrogen. Proteins, fusion proteins, or polypeptides can also be tagged with an epitope, such as a "Flag" epitope (Kodak), and purified using an antibody which specifically binds to that epitope.

The coding sequence disclosed herein can also be used to construct transgenic animals, such as mice, rats, guinea pigs, cows, goats, pigs, or sheep. Female transgenic animals can then produce proteins, polypeptides, or fusion proteins of the invention in their milk. Methods for constructing such animals are known and widely used in the art.

Alternatively, synthetic chemical methods, such as solid phase peptide synthesis, can be used to synthesize a secreted protein or polypeptide. General means for the production of peptides, analogs or derivatives are outlined in Chemistry and Biochemistry of Amino Acids, Peptides, and Proteins -- A Survey of Recent Developments, B. Weinstein, ed. (1983). Substitution of D-amino acids for the normal L-stereoisomer can be carried out to increase the half-life of the molecule.

Typically, homologous polynucleotide sequences can be confirmed by hybridization under stringent conditions, as is known in the art. For example, using the following wash conditions: 2 x SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.0), 0.1% SDS, room temperature twice, 30 minutes each; then 2 x SSC, 0.1% SDS, 50 °C once, 30 minutes; then 2 x SSC, room temperature twice, 10 minutes each, homologous sequences can be identified which contain at most about 25-30% basepair mismatches. More preferably, homologous nucleic acid strands contain 15-25% basepair mismatches, even more preferably 5-15% basepair mismatches.

The invention also provides polynucleotide probes which can be used to detect complementary nucleotide sequences, for example, in hybridization protocols such as Northern or Southern blotting or *in situ* hybridizations. Polynucleotide probes of the invention comprise at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, or 40 or more contiguous nucleotides of the gene A and gene B nucleic acid sequences provided herein. Polynucleotide probes of the invention can comprise a detectable label, such as a radioisotopic, fluorescent, enzymatic, or chemiluminescent label.

Isolated genes corresponding to the cDNA sequences disclosed herein are also provided. Standard molecular biology methods can be used to isolate the corresponding genes using the cDNA sequences provided herein. These methods include preparation of probes or primers from the nucleotide sequence shown in the figures for use in identifying or amplifying the genes from mammalian, including human, genomic libraries or other sources of human genomic DNA.

Polynucleotide molecules of the invention can also be used as primers to obtain additional copies of the polynucleotides, using polynucleotide amplification methods. Polynucleotide molecules can be propagated in vectors and cell lines using techniques well known in the art. Polynucleotide molecules can be on linear or circular molecules. They can be on autonomously replicating molecules or on molecules without replication sequences. They can be regulated by their own or by other regulatory sequences, as is known in the art.

Polynucleotide Constructs

Polynucleotide molecules comprising the coding sequences disclosed herein can be used in a polynucleotide construct, such as a DNA or RNA construct. Polynucleotide molecules of the invention can be used, for example, in an expression construct to express all or a portion of a protein, variant, fusion protein, or single-chain antibody in a host cell. An expression construct comprises a promoter which is functional in a chosen host cell. The skilled artisan can readily select an appropriate promoter from the large number of cell type-specific promoters known and used in the art. The expression construct can also contain a transcription terminator which is functional in the host cell. The expression construct comprises a polynucleotide segment which encodes all or a portion of the desired protein. The polynucleotide segment is located downstream from the promoter. Transcription of the polynucleotide segment initiates at the promoter. The expression construct can be linear or circular and can contain sequences, if desired, for autonomous replication.

Also included are polynucleotide molecules comprising human or non-human primate gene promoter and UTR sequences, operably linked to either protein coding sequences or other sequences encoding a detectable or selectable marker. Such promoter and/or UTR-based constructs are useful for studying the transcriptional and translational regulation of protein expression, and for identifying activating and/or inhibitory regulatory proteins.

Host Cells

An expression construct can be introduced into a host cell. The host cell comprising the expression construct can be any suitable prokaryotic or eukaryotic cell. Expression systems in bacteria include those described in Chang *et al., Nature* 275:615 (1978); Goeddel *et al., Nature* 281: 544 (1979); Goeddel *et al., Nucleic Acids Res.* 8:4057 (1980); EP 36,776; U.S. 4,551,433; deBoer *et al., Proc. Natl. Acad Sci. USA* 80: 21-25 (1983); and Siebenlist *et al., Cell* 20: 269 (1980).

Expression systems in yeast include those described in Hinnnen *et al., Proc. Natl. Acad. Sci. USA* 75: 1929 (1978); Ito *et al., J Bacteriol* 153: 163 (1983); Kurtz *et al., Mol. Cell. Biol.* 6: 142 (1986); Kunze *et al., J Basic Microbiol.* 25: 141 (1985); Gleeson *et al., J. Gen. Microbiol.* 132: 3459 (1986), Roggenkamp *et al., Mol. Gen. Genet.* 202: 302 (1986)); Das *et al.*, J Bacteriol. 158: 1165 (1984); De Louvencourt *et al., J Bacteriol.* 154:737 (1983), Van den Berg *et al., Bio/Technology* 8: 135 (1990); Kunze et al., *J. Basic Microbiol.* 25: 141 (1985); Cregg *et al., Mol. Cell. Biol.* 5: 3376 (1985); U.S. 4,837,148; U.S. 4,929,555; Beach and Nurse, *Nature* 300: 706 (1981); Davidow *et al., Curr. Genet.* 10: 380 (1985); Gaillardin *et al.*, Curr. *Genet.* 10: 49 (1985); Ballance *et al., Biochem. Biophys. Res. Commun.* 112: 284-289 (1983); Tilburn *et al., Gene* 26: 205-22 (1983); Yelton *et al., Proc. Natl. Acad, Sci. USA* 81: 1470-1474 (1984); Kelly and Hynes, *EMBO J.* 4: 475479 (1985); EP 244,234; and WO 91/00357.

Expression of heterologous genes in insects can be accomplished as described in U.S. 4,745,051; Friesen *et al.* (1986) "The Regulation of Baculovirus Gene Expression" in: THE MOLECULAR BIOLOGY OF BACULOVIRUSES (W. Doerfler, ed.); EP 127,839; EP 155,476; Vlak *et al., J. Gen. Virol.* 69: 765-776 (1988); Miller *et al., Ann. Rev. Microbiol.* 42: 177 (1988); Carbonell *et al., Gene* 73: 409 (1988); Maeda *et al., Nature* 315: 592-594 (1985); Lebacq-Verheyden *et al., Mol. Cell Biol.* 8: 3129 (1988); Smith *et al., Proc. Natl. Acad. Sci. USA* 82: 8404 (1985); Miyajima *et al., Gene* 58: 273 (1987); and Martin *et al., DNA* 7:99 (1988). Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts are described in Luckow *et al., Bio/Technology* (1988) 6: 47-55, Miller *et al.*, in GENETIC ENGINEERING (Setlow, J.K. *et al.* eds.), Vol. 8, pp. 277-279 (Plenum Publishing, 1986); and Maeda *et al., Nature,* 315: 592-594 (1985).

Mammalian expression can be accomplished as described in Dijkema *et al., EMBO J.* 4: 761(1985); *Gormanetal., Proc. Natl. Acad. Sci. USA* 79: 6777 (1982b); Boshart *et al., Cell*

41: 521 (1985); *and U.S.* 4,399,216. Other features of mammalian expression can be facilitated as described in Ham and Wallace, *Meth Enz.* 58: 44 (1979); Barnes and Sato, *Anal. Biochem.* 102: 255 (1980); U.S. 4,767,704; U.S. 4,657,866; U.S. 4,927,762; U.S. 4,560,655; WO 90/103430, WO 87/00195, and U.S. RE 30,985.

Expression constructs can be introduced into host cells using any technique known in the art. These techniques include transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and calcium phosphate-mediated transfection.

Expression of an endogenous gene encoding a protein of the invention can also be manipulated by introducing by homologous recombination a DNA construct comprising a transcription unit in frame with the endogenous gene, to form a homologously recombinant cell comprising the transcription unit. The transcription unit comprises a targeting sequence, a regulatory sequence, an exon, and an unpaired splice donor site. The new transcription unit can be used to turn the endogenous gene on or off as desired. This method of affecting endogenous gene expression is taught in U.S. Patent 5,641,670.

The targeting sequence is a segment of at least 10, 12, 15, 20, or 50 contiguous nucleotides of the nucleotide sequence shown in the figures herein. The transcription unit is located upstream to a coding sequence of the endogenous gene. The exogenous regulatory sequence directs transcription of the coding sequence of the endogenous gene.

Human or non-human primate protein can also include hybrid and modified forms thereof including fusion proteins, fragments and hybrid and modified forms in which certain amino acids have been deleted or replaced, modifications such as where one or more amino acids have been changed to a modified amino acid or unusual amino acid.

Also included within the meaning of substantially homologous is any human or non-human primate protein which may be isolated by virtue of cross-reactivity with antibodies to a gene described herein or whose encoding nucleotide sequences including genomic DNA, mRNA or cDNA may be isolated through hybridization with the complementary sequence of genomic or subgenomic nucleotide sequences or cDNA of a gene disclosed herein or a fragment thereof. It will also be appreciated by one skilled in the art that degenerate DNA sequences can encode human or non-human primate proteins and these are also intended to be included within the present invention as are allelic variants of.

Preferred is a colon or colorectal protein according to the invention prepared by recombinant DNA technology. By "pure form" or "purified form" or "substantially purified form" it is meant that a protein composition is substantially free of other proteins which are not protein.

The present invention also includes therapeutic or pharmaceutical compositions comprising human or non-human primate proteins, fragments or variants according to the invention in an effective amount for treating patients with disease, and a method comprising administering a therapeutically effective amount of a protein according to the invention. These compositions and methods are useful for treating cancers associated with a protein according to the invention, e.g. colon cancer. One skilled in the art can readily use a variety of assays known in the art to determine whether a protein according to the invention would be useful in promoting survival or functioning in a particular cell type.

In certain circumstances, it may be desirable to modulate or decrease the amount of the subject colon or colorectal protein expressed. Thus, in another aspect of the present invention, anti-sense oligonucleotides can be made specific to genes disclosed infra and a method utilized for diminishing the level of expression a protein according to the invention by a cell comprising administering one or more gene anti-sense oligonucleotides. By gene specific anti-sense oligonucleotides reference is made to oligonucleotides that have a nucleotide sequence that interacts through base pairing with a specific complementary nucleic acid sequence involved in the expression of a gene according to the invention that the expression of the gene is reduced. Preferably, the specific nucleic acid sequence involved in the expression of the subject gene is a genomic DNA molecule or mRNA molecule that encodes a colon or colorectal gene disclosed infra. This genomic DNA molecule can comprise regulatory regions of the gene, or the coding sequence for mature gene encoded by the gene.

The term complementary to a nucleotide sequence in the context of antisense oligonucleotides and methods therefor means sufficiently complementary to such a sequence as to allow hybridization to that sequence in a cell, *i.e.*, under physiological conditions. The antisense oligonucleotides preferably comprise a sequence containing from about 8 to about 100 nucleotides and more preferably the antisense oligonucleotides comprise from about 15 to about 30 nucleotides. The antisense oligonucleotides can also contain a variety of modifications that confer resistance to nucleolytic degradation such as, for example, modified

internucleoside linages [Uhlmann and Peyman, *Chemical Reviews* 90:543-548 (1990); Schneider and Banner, *Tetrahedron Lett.* 31:335, (1990) which are incorporated by reference], modified nucleic acid bases as disclosed in 5,958,773 and patents disclosed therein, and/or sugars and the like.

Any modifications or variations of the antisense molecule which are known in the art to be broadly applicable to antisense technology are included within the scope of the invention. Such modifications include preparation of phosphorus-containing linkages as disclosed in U.S. Patents 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361, 5,625,050 and 5,958,773.

The antisense compounds of the invention can include modified bases. The antisense oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, cellular distribution, or cellular uptake of the antisense oligonucleotide. Such moieties or conjugates include lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, U.S. Patents 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773.

Chimeric antisense oligonucleotides are also within the scope of the invention, and can be prepared from the present inventive oligonucleotides using the methods described in, for example, U.S. Patents 5,013,830, 5,149,797, 5,403,711, 5,491,133, 5,565,350, 5,652,355, 5,700,922 and 5,958,773.

In the antisense art a certain degree of routine experimentation is required to select optimal antisense molecules for particular targets. To be effective, the antisense molecule preferably is targeted to an accessible, or exposed, portion of the target RNA molecule. Although in some cases information is available about the structure of target mRNA molecules, the current approach to inhibition using antisense is via experimentation. mRNA levels in the cell can be measured routinely in treated and control cells by reverse transcription of the mRNA and assaying the cDNA levels. The biological effect can be determined routinely by measuring cell growth or viability as is known in the art.

Measuring the specificity of antisense activity by assaying and analyzing cDNA levels is an art-recognized method of validating antisense results. It has been suggested that RNA

from treated and control cells should be reverse-transcribed and the resulting cDNA populations analyzed. [Branch, A. D., *T.I.B.S.* 23:45-50 (1998)].

The therapeutic or pharmaceutical compositions of the present invention can be administered by any suitable route known in the art including for example intravenous, subcutaneous, intramuscular, transdermal, intrathecal or intracerebral. Administration can be either rapid as by injection or over a period of time as by slow infusion or administration of slow release formulation.

Additionally, a human or non-human primate protein according to the invention can also be linked or conjugated with agents that provide desirable pharmaceutical or pharmacodynamic properties. For example, the protein can be coupled to any substance known in the art to promote penetration or transport across the blood-brain barrier such as an antibody to the transferrin receptor, and administered by intravenous injection (see, for example, Friden et al., *Science* 259:373-377 (1993) which is incorporated by reference). Furthermore, the subject protein can be stably linked to a polymer such as polyethylene glycol to obtain desirable properties of solubility, stability, half-life and other pharmaceutically advantageous properties. [See, for example, Davis et al., *Enzyme Eng.* 4:169-73 (1978); Buruham, *Am. J. Hosp. Pharm.* 51:210-218 (1994) which are incorporated by reference].

The compositions are usually employed in the form of pharmaceutical preparations. Such preparations are made in a manner well known in the pharmaceutical art. See, e.g. Remington Pharmaceutical Science, 18th Ed., Merck Publishing Co. Eastern PA, (1990). One preferred preparation utilizes a vehicle of physiological saline solution, but it is contemplated that other pharmaceutically acceptable carriers such as physiological concentrations of other non-toxic salts, five percent aqueous glucose solution, sterile water or the like may also be used. It may also be desirable that a suitable buffer be present in the composition. Such solutions can, if desired, be lyophilized and stored in a sterile ampoule ready for reconstitution by the addition of sterile water for ready injection. The primary solvent can be aqueous or alternatively non-aqueous. The subject human or primate protein, fragment or variant thereof can also be incorporated into a solid or semi-solid biologically compatible matrix which can be implanted into tissues requiring treatment.

The carrier can also contain other pharmaceutically-acceptable excipients for modifying or maintaining the pH, osmolarity, viscosity, clarity, color, sterility, stability, rate of dissolution, or odor of the formulation. Similarly, the carrier may contain still other

pharmaceutically-acceptable excipients for modifying or maintaining release or absorption or penetration across the blood-brain barrier. Such excipients are those substances usually and customarily employed to formulate dosages for parenteral administration in either unit dosage or multi-dose form or for direct infusion into the cerebrospinal fluid by continuous or periodic infusion.

Dose administration can be repeated depending upon the pharmacokinetic parameters of the dosage formulation and the route of administration used.

It is also contemplated that certain formulations containing a protein according to the invention or variant or fragment thereof are to be administered orally. Such formulations are preferably encapsulated and formulated with suitable carriers in solid dosage forms. Some examples of suitable carriers, excipients, and diluents include lactose, dextrose, sucrose, sorbitol, mannitol, starches, gum acacia, calcium phosphate, alginates, calcium silicate, microcrystalline cellulose, polyvinylpyrrolidone, cellulose, gelatin, syrup, methyl cellulose, methyl- and propylhydroxybenzoates, talc, magnesium, stearate, water, mineral oil, and the like. The formulations can additionally include lubricating agents, wetting agents, emulsifying and suspending agents, preserving agents, sweetening agents or flavoring agents. The compositions may be formulated so as to provide rapid, sustained, or delayed release of the active ingredients after administration to the patient by employing procedures well known in the art. The formulations can also contain substances that diminish proteolytic degradation and promote absorption such as, for example, surface active agents.

The specific dose is calculated according to the approximate body weight or body surface area of the patient or the volume of body space to be occupied. The dose will also be calculated dependent upon the particular route of administration selected. Further refinement of the calculations necessary to determine the appropriate dosage for treatment is routinely made by those of ordinary skill in the art. Such calculations can be made without undue experimentation by one skilled in the art in light of the activity disclosed herein in assay preparations of target cells. Exact dosages are determined in conjunction with standard dose-response studies. It will be understood that the amount of the composition actually administered will be determined by a practitioner, in the light of the relevant circumstances including the condition or conditions to be treated, the choice of composition to be administered, the age, weight, and response of the individual patient, the severity of the patient's symptoms, and the chosen route of administration.

In one embodiment of this invention, a protein according to the invention may be therapeutically administered by implanting into patients vectors or cells capable of producing a biologically-active form of the protein or a precursor of the protein, *i.e.*, a molecule that can be readily converted to a biological-active form of the by the body. In one approach, cells that secrete the protein may be encapsulated into semipermeable membranes for implantation into a patient. The cells can be cells that normally express the protein or a precursor thereof or the cells can be transformed to express the protein or a precursor thereof. It is preferred that the cell be of human origin and that the protein comprise the native human protein when the patient is human. However, it is anticipated that a non-human primate protein homolog of a human protein according to the invention may be effective.

In a number of circumstances it would be desirable to determine the levels of protein or corresponding mRNA encoding a protein according to the invention in a patient. The identification of the subject genes which are specifically expressed by colon or colorectal tumors suggests these proteins may be expressed at different levels during some diseases, e.g., cancers, provides the basis for the conclusion that the presence of these proteins serves a normal physiological function related to cell growth and survival. Endogenously produced human colon or colorectal antigen according to the invention may also play a role in certain disease conditions.

The term "detection" as used herein in the context of detecting the presence of a cancer gene according to the invention in a patient is intended to include the determining of the amount of protein according to the invention or the ability to express an amount of this protein in a patient, the estimation of prognosis in terms of probable outcome of a disease and prospect for recovery, the monitoring of these protein levels over a period of time as a measure of status of the condition, and the monitoring of colon or colorectal protein according to the invention for determining a preferred therapeutic regimen for the patient, e.g. one with colon cancer.

To detect the presence of a gene according to the invention in a patient, a sample is obtained from the patient. The sample can be a tissue biopsy sample or a sample of blood, plasma, serum, CSF or the like. It has been found that the subject genes are expressed at high levels in some cancers, e.g., colon or colorectal cancers. Samples for detecting protein can be taken from these tissue. When assessing peripheral levels of protein, it is preferred that the sample be a sample of blood, plasma or serum. When assessing the levels of protein in the

central nervous system a preferred sample is a sample obtained from cerebrospinal fluid or neural tissue.

In some instances, it is desirable to determine whether a gene according to the invention is intact in the patient or in a tissue or cell line within the patient. By an intact gene, it is meant that there are no alterations in the gene such as point mutations, deletions, insertions, chromosomal breakage, chromosomal rearrangements and the like wherein such alteration might alter the production of gene or alter its biological activity, stability or the like to lead to disease processes. Thus, in one embodiment of the present invention a method is provided for detecting and characterizing any alterations in the gene. The method comprises providing an oligonucleotide that contains the gene corresponding cDNA, genomic DNA or a fragment thereof or a derivative thereof. By a derivative of an oligonucleotide, it is meant that the derived oligonucleotide is substantially the same as the sequence from which it is derived in that the derived sequence has sufficient sequence complementarily to the sequence from which it is derived to hybridize specifically to the gene. The derived nucleotide sequence is not necessarily physically derived from the nucleotide sequence, but may be generated in any manner including for example, chemical synthesis or DNA replication or reverse transcription or transcription.

Typically, patient genomic DNA is isolated from a cell sample from the patient and digested with one or more restriction endonucleases such as, for example, TaqI and AluI. Using the Southern blot protocol, which is well known in the art, this assay determines whether a patient or a particular tissue in a patient has an intact gene according to the invention or a gene abnormality.

Hybridization to a gene according to the invention would involve denaturing the chromosomal DNA to obtain a single-stranded DNA; contacting the single-stranded DNA with a gene probe associated with the gene sequence; and identifying the hybridized DNA-probe to detect chromosomal DNA containing at least a portion of a human gene according to the invention.

The term "probe" as used herein refers to a structure comprised of a polynucleotide that forms a hybrid structure with a target sequence, due to complementarity of probe sequence with a sequence in the target region. Oligomers suitable for use as probes may contain a minimum of about 8-12 contiguous nucleotides which are complementary to the targeted sequence and preferably a minimum of about 20.

Probes of the present invention can be DNA or RNA oligonucleotides and can be made by any method known in the art such as, for example, excision, transcription or chemical synthesis. Probes may be labeled with any detectable label known in the art such as, for example, radioactive or fluorescent labels or enzymatic marker. Labeling of the probe can be accomplished by any method known in the art such as by PCR, random priming, end labeling, nick translation or the like. One skilled in the art will also recognize that other methods not employing a labeled probe can be used to determine the hybridization. Examples of methods that can be used for detecting hybridization include Southern blotting, fluorescence in situ hybridization, and single-strand conformation polymorphism with PCR amplification.

Hybridization is typically carried out at 25° - 45° C, more preferably at 32° -40° C and more preferably at 37° - 38° C. The time required for hybridization is from about 0.25 to about 96 hours, more preferably from about one to about 72 hours, and most preferably from about 4 to about 24 hours.

Gene abnormalities can also be detected by using the PCR method and primers that flank or lie within the particular gene. The PCR method is well known in the art. Briefly, this method is performed using two oligonucleotide primers which are capable of hybridizing to the nucleic acid sequences flanking a target sequence that lies within gene and amplifying the target sequence. The terms "oligonucleotide primer" as used herein refers to a short strand of DNA or RNA ranging in length from about 8 to about 30 bases. The upstream and downstream primers are typically from about 20 to about 30 base pairs in length and hybridize to the flanking regions for replication of the nucleotide sequence. The polymerization is catalyzed by a DNA-polymerase in the presence of deoxynucleotide triphosphates or nucleotide analogs to produce double-stranded DNA molecules. The double strands are then separated by any denaturing method including physical, chemical or enzymatic. Commonly, a method of physical denaturation is used involving heating the nucleic acid, typically to temperatures from about 80°C to 105°C for times ranging from about 1 to about 10 minutes. The process is repeated for the desired number of cycles.

The primers are selected to be substantially complementary to the strand of DNA being amplified. Therefore, the primers need not reflect the exact sequence of the template, but must be sufficiently complementary to selectively hybridize with the strand being amplified.

After PCR amplification, the DNA sequence comprising a gene of the invention or a fragment thereof is then directly sequenced and analyzed by comparison of the sequence with the sequences disclosed herein to identify alterations which might change activity or expression levels or the like.

In another embodiment, a method for detecting protein a colon according to the invention is provided based upon an analysis of tissue expressing the gene. Certain tissues such as breast, lung, colon and others maybe analyzed. The method comprises hybridizing a polynucleotide to mRNA from a sample of tissue that normally expresses the gene. The sample is obtained from a patient suspected of having an abnormality in the gene.

To detect the presence of mRNA encoding protein a colon or colorectal protein according to the invention is obtained from a patient. The sample can be from blood or from a tissue biopsy sample. The sample may be treated to extract the nucleic acids contained therein. The resulting nucleic acid from the sample is subjected to gel electrophoresis or other size separation techniques.

The mRNA of the sample is contacted with a DNA sequence serving as a probe to form hybrid duplexes. The use of a labeled probes as discussed above allows detection of the resulting duplex.

When using the cDNA encoding a colon or colorectal protein according to the invention or a derivative of the cDNA as a probe, high stringency conditions can be used in order to prevent false positives, that is the hybridization and apparent detection of the gene nucleotide sequences when in fact an intact and functioning gene is not present. When using sequences derived from the gene or cDNA, less stringent conditions could be used, however, this would be a less preferred approach because of the likelihood of false positives. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. [Sambrook et al. (1989), *supra*].

In order to increase the sensitivity of the detection in a sample of mRNA encoding the protein, the technique of reverse transcription/ polymerization chain reaction (RT/PCR) can be used to amplify cDNA transcribed from mRNA encoding the protein. The method of RT/PCR is well known in the art, and can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is

reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified using the PCR method and specific primers. [Belyavsky et al., *Nucl. Acid Res.* 17:2919-2932 (1989); Krug and Berger, *Methods in Enzymology*, 152:316-325, Academic Press, NY (1987) which are incorporated by reference].

The polymerase chain reaction method is performed as described above using two oligonucleotide primers that are substantially complementary to the two flanking regions of the DNA segment to be amplified. Following amplification, the PCR product is then electrophoresed and detected by ethidium bromide staining or by phosphoimaging.

The present invention further provides for methods to detect the presence of a colon or colorectal protein in a sample obtained from a patient. Any method known in the art for detecting proteins can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. [*Basic and Clinical Immunology*, 217-262, Sites and Terr, eds., Appleton & Lange, Norwalk, CT, (1991), which is incorporated by reference]. Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes of a colon protein according to the invention and competitively displacing a labeled protein or derivative thereof.

As used herein, a derivative of a protein according to the invention is intended to include a polypeptide in which certain amino acids have been deleted or replaced or changed to modified or unusual amino acids wherein the derivative is biologically equivalent to the gene and wherein the polypeptide derivative cross-reacts with antibodies raised against the protein. By cross-reaction it is meant that an antibody reacts with an antigen other than the one that induced its formation.

Numerous competitive and non-competitive protein-binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labeled for use in a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like for use in radioinununoassay (RIA), enzyme immunoassays, *e.g.*, enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like.

Polyclonal or monoclonal antibodies to the subject non-human primate or human proteins or according to the invention an epitope thereof can be made for use in immunoassays by any of a number of methods known in the art. By epitope reference is made to an antigenic determinant of a polypeptide. An epitope could comprise 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope consists of at least 5 such amino acids. Methods of determining the spatial conformation of amino acids are known in the art, and include, for example, x-ray crystallography and 2 dimensional nuclear magnetic resonance.

One approach for preparing antibodies to a protein is the selection and preparation of an amino acid sequence of all or part of the protein, chemically synthesizing the sequence and injecting it into an appropriate animal, typically a rabbit, hamster or a mouse.

Oligopeptides can be selected as candidates for the production of an antibody to the subject colon or colorectal protein based upon the oligopeptides lying in hydrophilic regions, which are thus likely to be exposed in the mature protein.

Additional oligopeptides can be determined using, for example, the Antigenicity Index, Welling, G.W. et al., *FEBS Lett.* 188:215-218 (1985), incorporated herein by reference.

In other embodiments of the present invention, humanized monoclonal antibodies are provided, wherein the antibodies are specific for a protein according to the invention. The phrase "humanized antibody" refers to an antibody derived from a non-human antibody, typically a mouse monoclonal antibody. Alternatively, a humanized antibody may be derived from a chimeric antibody that retains or substantially retains the antigen-binding properties of the parental, non-human, antibody but which exhibits diminished immunogenicity as compared to the parental antibody when administered to humans. The phrase "chimeric antibody," as used herein, refers to an antibody containing sequence derived from two different antibodies (*see, e.g.,* U.S. Patent No. 4,816,567) which typically originate from different species. Most typically, chimeric antibodies comprise human and murine antibody fragments generally human constant and mouse variable regions.

Because humanized antibodies are far less immunogenic in humans than the parental mouse monoclonal antibodies, they can be used for the treatment of humans with far less risk of anaphylaxis. Thus, these antibodies may be preferred in therapeutic applications that involve *in vivo* administration to a human such as, *e.g.,* use as radiation sensitizers for the

treatment of neoplastic disease or use in methods to reduce the side effects of, *e.g.*, cancer therapy.

Humanized antibodies may be achieved by a variety of methods including, for example: (1) grafting the non-human complementarity determining regions (CDRs) onto a human framework and constant region (a process referred to in the art as "humanizing"), or, alternatively, (2) transplanting the entire non-human variable domains, but "cloaking" them with a human-like surface by replacement of surface residues (a process referred to in the art as "veneering"). In the present invention, humanized antibodies will include both "humanized" and "veneered" antibodies. These methods are disclosed in, *e.g.*, Jones et al., *Nature* 321:522-525 (1986); Morrison et al., *Proc. Natl. Acad. Sci, US.A.*, 81:6851-6855 (1984); Morrison and Oi, *Adv. Immunol.*, 44:65-92 (1988); Verhoeyer et al., *Science 239*:1534-1536 (1988); Padlan, *Molec. Immun.* 28:489-498 (1991); Padlan, *Molec. Immunol.* 31(3): 169-217 (1994); and Kettleborough, C.A. et al., *Protein Eng.* 4(7):773-83 (1991) each of which is incorporated herein by reference.

The phrase "complementarity determining region" refers to amino acid sequences which together define the binding affinity and specificity of the natural Fv region of a native immunoglobulin-binding site. *See, e.g.*, Chothia et al., *J. Mol. Biol.* 196:901-917 (1987); Kabat et al., U.S. Dept. of Health and Human Services NIH Publication No. 91-3242 (1991). The phrase "constant region" refers to the portion of the antibody molecule that confers effector functions. In the present invention, mouse constant regions are substituted by human constant regions. The constant regions of the subject-humanized antibodies are derived from human immunoglobulins. The heavy chain constant region can be selected from any of the five isotypes: alpha, delta, epsilon, gamma or mu.

One method of humanizing antibodies comprises aligning the non-human heavy and light chain sequences to human heavy and light chain sequences, selecting and replacing the non-human framework with a human framework based on such alignment, molecular modeling to predict the conformation of the humanized sequence and comparing to the conformation of the parent antibody. This process is followed by repeated back mutation of residues in the CDR region which disturb the structure of the CDRs until the predicted conformation of the humanized sequence model closely approximates the conformation of the non-human CDRs of the parent non-human antibody. Such humanized antibodies may be

further derivatized to facilitate uptake and clearance, *e.g.* via Ashwell receptors. *See, e.g.*, U.S. Patent Nos. 5,530,101 and 5,585,089 which patents are incorporated herein by reference.

Humanized antibodies to proteins according to the invention can also be produced using transgenic animals that are engineered to contain human immunoglobulin loci. For example, WO 98/24893 discloses transgenic animals having a human Ig locus wherein the animals do not produce functional endogenous immunoglobulins due to the inactivation of endogenous heavy and light chain loci. WO 91/10741 also discloses transgenic non-primate mammalian hosts capable of mounting an immune response to an immunogen, wherein the antibodies have primate constant and/or variable regions, and wherein the endogenous immunoglobulin-encoding loci are substituted or inactivated. WO 96/30498 discloses the use of the Cre/Lox system to modify the immunoglobulin locus in a mammal, such as to replace all or a portion of the constant or variable region to form a modified antibody molecule. WO 94/02602 discloses non-human mammalian hosts having inactivated endogenous Ig loci and functional human Ig loci. U.S. Patent No. 5,939,598 discloses methods of making transgenic mice in which the mice lack endogenous heavy claims, and express an exogenous immunoglobulin locus comprising one or more xenogeneic constant regions.

Using a transgenic animal described above, an immune response can be produced to a selected antigenic molecule, and antibody-producing cells can be removed from the animal and used to produce hybridomas that secrete human monoclonal antibodies. Immunization protocols, adjuvants, and the like are known in the art, and are used in immunization of, for example, a transgenic mouse as described in WO 96/33735. This publication discloses monoclonal antibodies against a variety of antigenic molecules including IL-6, IL-8, TNF, human CD4, L-selectin, gp39, and tetanus toxin. The monoclonal antibodies can be tested for the ability to inhibit or neutralize the biological activity or physiological effect of the corresponding protein. WO 96/33735 discloses that monoclonal antibodies against IL-8, derived from immune cells of transgenic mice immunized with IL-8, blocked IL-8-induced functions of neutrophils. Human monoclonal antibodies with specificity for the antigen used to immunize transgenic animals are also disclosed in WO 96/34096.

In the present invention, proteins and variants thereof according to the invention are used to immunize a transgenic animal as described above. Monoclonal antibodies are made using methods known in the art, and the specificity of the antibodies is tested using isolated colon or colorectal proteins according to the invention.

Methods for preparation of the human or primate protein according to the invention or an epitope thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples. Chemical synthesis of a peptide can be performed, for example, by the classical Merrifeld method of solid phase peptide synthesis (Merrifeld, *J. Am. Chem. Soc. 85*:2149, 1963 which is incorporated by reference) or the FMOC strategy on a Rapid Automated Multiple Peptide Synthesis system [E. I. du Pont de Nemours Company, Wilmington, DE) (Caprino and Han, *J. Org. Chem.* 37:3404 (1972) which is incorporated by reference].

Polyclonal antibodies can be prepared by immunizing rabbits or other animals by injecting antigen followed by subsequent boosts at appropriate intervals. The animals are bled and sera assayed against purified protein usually by ELISA or by bioassay based upon the ability to block the action of a gene according to the invention. When using avian species, e.g., chicken, turkey and the like, the antibody can be isolated from the yolk of the egg. Monoclonal antibodies can be prepared after the method of Milstein and Kohler by fusing splenocytes from immunized mice with continuously replicating tumor cells such as myeloma or lymphoma cells. [Milstein and Kohler, *Nature 256*:495-497 (1975); Gulfre and Milstein, *Methods in Enzymology: Immunochemical Techniques 73*:1-46, Langone and Banatis eds., Academic Press, (1981) which are incorporated by reference]. The hybridoma cells so formed are then cloned by limiting dilution methods and supernates assayed for antibody production by ELISA, RIA or bioassay.

The unique ability of antibodies to recognize and specifically bind to target proteins provides an approach for treating an overexpression of the protein. Thus, another aspect of the present invention provides for a method for preventing or treating diseases involving overexpression of the a protein according to the invention by treatment of a patient with antibodies to specific tumor antigen according to the invention.

Specific antibodies, either polyclonal or monoclonal, to the protein can be produced by any suitable method known in the art as discussed above. For example, murine or human monoclonal antibodies can be produced by hybridoma technology or, alternatively, the tumor protein, or an immunologically active fragment thereof, or an anti-idiotypic antibody, or fragment thereof can be administered to an animal to elicit the production of antibodies capable of recognizing and binding to the tumor protein. Such antibodies can be from any

class of antibodies including, but not limited to IgG, IgA, 1gM, IgD, and IgE or in the case of avian species, IgY and from any subclass of antibodies.

The availability of isolated human or primate protein according to the invention allows for the identification of small molecules and low molecular weight compounds that inhibit the binding of the protein to binding partners, through routine application of high-throughput screening methods (HTS). HTS methods generally refer to technologies that permit the rapid assaying of lead compounds for therapeutic potential. HTS techniques employ robotic handling of test materials, detection of positive signals, and interpretation of data. Lead compounds may be identified via the incorporation of radioactivity or through optical assays that rely on absorbance, fluorescence or luminescence as read-outs. [Gonzalez, J.E. *et al., Curr. Opin. Biotech.* 9:624-63 1 (1998)].

Model systems are available that can be adapted for use in high throughput screening for compounds that inhibit the interaction of a protein with its ligand, for example by competing with the protein for ligand binding. Sarubbi *et al., Anal. Biochem. 237*:70-75 (1996) describe cell-free, non-isotopic assays for discovering molecules that compete with natural ligands for binding to the active site of IL-1 receptor. Martens, C. *et al., Anal. Biochem. 273*:20-31 (1999) describe a generic particle-based nonradioactive method in which a labeled ligand binds to its receptor immobilized on a particle; label on the particle decreases in the presence of a molecule that competes with the labeled ligand for receptor binding.

The therapeutic gene polynucleotides and polypeptides of the present invention may be utilized in gene delivery vehicles. The gene delivery vehicle may be of viral or non-viral origin (*see generally*, Jolly, *Cancer Gene Therapy* 1:51-64 (1994); Kimura, *Human Gene Therapy* 5:845-852 (1994); Connelly, *Human Gene Therapy* 1:185-193 (1995); and Kaplitt, *Nature Genetics* 6:148-153 (1994)). Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic according to the invention can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches. Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

The present invention can employ recombinant retroviruses which are constructed to carry or express a selected nucleic acid molecule of interest. Retrovirus vectors that can be employed include those described in EP 0 415 731; WO 90/07936; WO 94/03622; WO

93/25698; WO 93/25234; U.S. Patent No. 5,219,740; WO 93/11230; WO 93/10218; Vile and Hart, *Cancer Res.* 53:3860-3864 (1993); Vile and Hart, *Cancer Res.* 53:962-967 (1993); Ram et al., *Cancer Res.* 53:83-88 (1993); Takamiya et al., *J. Neurosci. Res.* 33:493-503 (1992); Baba et al., *J. Neurosurg.* 79:729-735 (1993); U.S. Patent No. 4,777,127; GB Patent No. 2,200,651; and EP 0 345 242. Preferred recombinant retroviruses include those described in WO 91/02805.

Packaging cell lines suitable for use with the above-described retroviral vector constructs may be readily prepared (see PCT publications WO 95/3 0763 and WO 92/05266), and used to create producer cell lines (also termed vector cell lines) for the production of recombinant vector particles. Within particularly preferred embodiments of the invention, packaging cell lines are made from human (such as HT1080 cells) or mink parent cell lines, thereby allowing production of recombinant retroviruses that can survive inactivation in human serum.

The present invention also employs alphavirus-based vectors that can function as gene delivery vehicles. Such vectors can be constructed from a wide variety of alphaviruses, including, for example, Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532). Representative examples of such vector systems include those described in U.S. Patent Nos. 5,091,309; 5,217,879; and 5,185,440; and PCT Publication Nos. WO 92/10578; WO 94/21792; WO 95/27069; WO 95/27044; and WO 95/07994.

Gene delivery vehicles of the present invention can also employ parvovirus such as adeno-associated virus (AAV) vectors. Representative examples include the AAV vectors disclosed by Srivastava in WO 93/09239, Samulski et al., *J. Vir.* 63: 3822-3828 (1989); Mendelson et al., *Virol.* 166: 154-165 (1988); and Flotte et al., *P.N.A.S.* 90: 10613-10617 (1993).

Representative examples of adenoviral vectors include those described by Berkner, *Biotechniques 6*:616-627 (Biotechniques); Rosenfeld *et al., Science 252*:431-434 (1991); WO 93/19191; Kolls *et al., P.N.A.S.* 215-219 (1994); Kass-Bisleret al., P.N.A.S. 90: 11498-11502 (1993); Guzman et al., *Circulation* 88: 2838-2848 (1993); Guzman *et al., Cir. Res.* 73: 1202-1207 (1993); Zabner et al., *Cell* 75: 207-216 (1993); Li et al., *Hum. Gene Ther.* 4: 403-409 (1993); Cailaud et al., *Eur. J. Neurosci.* 5: 1287-1291 (1993); Vincent et al., *Nat. Genet.*

5: 130-134 (1993); Jaffe et al., *Nat. Genet.* 1: 372-378 (1992); and Levrero et al., *Gene* 101: 195-202 (1992). Exemplary adenoviral gene therapy vectors employable in this invention also include those described in WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655. Administration of DNA linked to kill adenovirus as described in Curiel, *Hum. Gene Ther.* 3: 147-154 (1992) may be employed.

Other gene delivery vehicles and methods may be employed, including polycationic condensed DNA linked or unlinked to kill adenovirus alone, for example Curiel, *Hum. Gene Ther.* 3: 147-154 (1992); ligand-linked DNA, for example see Wu, *J. Biol. Chem.* 264: 16985-16987 (1989); eukaryotic cell delivery vehicles cells, for example see U.S. Serial No. 08/240,030, filed May 9, 1994, and U.S. Serial No. 08/404,796; deposition of photopolymerized hydrogel materials; hand-held gene transfer particle gun, as described in U.S. Patent No. 5,149,655; ionizing radiation as described in U.S. Patent No. 5,206,152 and in WO 92/11033; nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip, *Mol. Cell Biol. 14*:2411-2418 (1994), and in Woffendin, *Proc. Natl. Acad. Sci. 91*:1581-1585 (1994).

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and U.S. Patent No. 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm. Liposomes that can act as gene delivery vehicles are described in U.S. Patent No. 5,422,120, PCT Patent Publication Nos. WO 95/13 796, WO 94/23697, and WO 9 1/14445, and EP No. 0 524 968.

Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin et al., *Proc. Natl. Acad. Sci. USA* 91(24): 11581-11585 (1994). Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in U.S. Patent No. 5,149,655; use of ionizing radiation for activating transferred gene, as described in U.S. Patent No. 5,206,152 and PCT Patent Publication No. WO 92/11033.

While the invention has been described supra, including preferred embodiments, the following examples are provided to further illustrate the invention.

EXAMPLE 1

Through a collaboration with Analytical Pathology Medical Group (at Grossmont Hospital) IDEC obtains pairs of snap frozen normal and malignant colon tissue obtained during surgery. RNA is extracted from 10 pairs of those samples and submitted for GeneTag analysis at Celera/Applied Bio Systems (ABI). In short, the RNA is reverse transcribed into cDNA, digested with a restriction enzyme, and linkers are ligated to the cDNA library. The library is amplified using the linker sequences as a primer with an additional nucleotide (A, T, G, or C) (+1 PCR) generating 16 libraries. These 16 libraries are further amplified using the linker sequences as primers with an additional two nucleotides (+2 PCR) generating 256 libraries. Fluorescently labeled products from these +2 PCR reactions are separated by capillary electrophoresis and the peaks are quantitated. We compared peaks obtained from the malignant colon RNA to peaks obtained using RNA from the normal colon and found a number that were five-fold overexpressed in three of three tumors. These peaks are purified and amplified by PCR using the linkers with three additional nucleotides (+3 PCR). The +3 peaks are purified and sequenced. These sequences are set forth below.

CICO1 Celera IDEC Colon Overexpressed 1 (CICO1)(bs213ms134-185)

Using 185 bases of +3 PCR sequence from Celera, we identified human tentative human consensus sequence (THC) 684921 from the BLAT database.

bs213ms143-185
GATCCAGGAGAGGAAGGAGTTTCAGAAGGCAGGAGCTGGTCCTCTATGTCATGAAATGTAGA
GGGTGAGGCCAAGGAGGACCTGAGAGAAGGTAATTAGATTTGGTGTTTACAGGCTGGTCCCT
GTGGCCAGCCACCCCACCCACTTTA(SEQ ID NO:1)

THC 684921
TGAGGAAACTGTGGCTTAGAGGAAAAGGTCATTAGTTCATTTTGGGATTT
GTTGATTTTCAGATGTTTGAGATGTTGAGGATGGATTGTCCAGCAGGCTA
TTAAGATGTGGTGAAGGCTAGAAATGTTGATTTAGGAGGTATTGCCTTCG

AGAAGATAAAGGAGGAGAAGAGGAGAGCATCATGCAAGCTAGAGAAGAGA
AAGAAGAAAAGTATTCTGGGGAATGTCTCCTTTGGGAGCAGAAAGAAGAC
TCTGACGGAGCAGCCATCCAGGAAGTGGAATGAGATCCAGGAGAGGAAGG
AGTTTCAGAAGGCAGGAGCTGGTCCTCTATGTCATGAAATGTAGAGGGTG
AGGCCAAGGAGGACCTGAGAGAAGGTAATTAGATTTGGTGTTTACAGGCT
GGTCCCTGTGGCCAGCCACCCCACCCACTTTAAAATATTTACTCTACAAA
TGTTAATGTGTGAAGAGTTGCATGCCAGAATATTTATGGCATCAGTGTTG
GTGGATACAGAACATTGGGAAACAACCCATTAATAGCAGAATGGTAAATC
TGGCCAGTGAATAGTATAGCTTTTTAAAAGGAGGCTGATGTCTGAATTCA
CTTTCAAAGTTGTTCACAATGTATTGCTAAAATACAAAAATGTTGCAGAA
CCATATGTATGAGAGAAACCCCTTTTTCT(SEQ ID NO:2)

CICO 2 (bs222ms233-191)

191 bases of the +3 PCR sequence from Celera overlapped with the 3'UTR of four different hypothetical proteins in the BLAT database.

bs222ms233-191
GATCCCCATGGTATGCTTGAATCTGCTCCCTGAACTTCCTGCCAGTGCCTCCCCGTACCCCA
AAACAATGTCACCATGGTTACCACCTACCCAGAAGACTGTTCCCTCCTCCCAAGACCCTTGT
CTGCAGTGGTGCTCCTGCAGGCTGCCCGTTA(SEQ ID NO:3)
chr1_70_2399.c mRNA sequence (coding in CAPITALS, no ATG at start)
AGTGTGGTGATGGTTGTCTTCGACAATGAGAAGGTCCCAGTAGAGCAGCT
GCGCTTCTGGAAGCACTGGCATTCCCGGCAACCCACTGCCAAGCAGCGGG
TCATTGACGTGGCTGACTGCAAAGAAAACTTCAACACTGTGGAGCACATT
GAGGAGGTGGCCTATAATGCACTGTCCTTTGTGTGGAACGTGAATGAAGA
GGCCAAGGTGTTCATCGGCGTAAACTGTCTGAGCACAGACTTTTCCTCAC
AAAAGGGGGTGAAGGGTGTCCCCCTGAACCTGCAGATTGACACCTATGAC
TGTGGCTTGGGCACTGAGCGCCTGGTACACCGTGCTGTCTGCCAGATCAA
GATCTTCTGTGACAAGGGAGCTGAGAGGAAGATGCGCGATGACGAGCGGA
AGCAGTTCCGGAGGAAGGTCAAGTGCCCTGACTCCAGCAACAGTGGCGTC
AAGGGCTGCCTGCTGTCGGGCTTCAGGGGCAATGAGACGACCTACCTTCG
GCCAGAGACTGACCTGGAGACGCCACCCGTGCTGTTCATCCCCAATGTGC
ACTTCTCCAGCCTGCAGCGGTCTGGAGGGGCAGCCCCCTCGGCAGGACCC
AGCAGCTCCAACAGGCTGCCTCTGAAGCGTACCTGCTCGCCCTTCACTGA

GGAGTTTGAGCCTCTGCCCTCCAAGCAGGCCAAGGAAGGCGACCTTCAGA
GAGTTCTGCTGTATGTGCGGAGGGAGACTGAGGAGGTGTTTGACGCGCTC
ATGTTGAAGACCCCAGACCTGAAGGGGCTGAGGAATGCGATCTCTGAGAA
GTATGGGTTCCCTGAAGAGAACATTTACAAAGTCTACAAGAAATGCAAGC
GAGGAATCTTAGTCAACATGGACAACAACATCATTCAGCATTACAGCAAC
CACGTCGCCTTCCTGCTGGACATGGGGGAGCTGGACGGCAAAATTCAGAT
CATCCTTAAGGAGCTGTAAggcctctcgagcatccaaaccctcacgacct
gcaaggggccagcagggacgtggccccacgccacacacaacctctccaca
tgcctcagcgctgttacttgaatgccttccctgagggaagaggcccttga
gtcacagacccacagacgtcagggccagggagagacctagggggtcccct
ggcctggatccccatggtatgcttgaatctgctccctgaacttcctgcca
gtgcctccccgtaccccaaaacaatgtcaccatggttaccacctacccag
aagactgttccctcctcccaagacccttgtctgcagtggtgctcctgcag
gctgcccgttaagatggtggcggcacacgctccctcccgcagcaccacgc
cagctggtgcggcccccactctctgtcttccttcaacttcagacaaagga
tttctcaacctttggtcagttaacttgaaaactcttgattttcagtgcaa
atgacttttaaaagacactatattggagtctctttctcagacttcctcag
cgcaggatgtaaatagcactaacgatcgactggaacaaagtgaccgctgt
gtaaaactactgccttgccactcactgttgtatacatttcttatttacga
ttttcatttgttatatatatatataaatatactgtatatatatgcaacat
tttatattttcatggatatgtttttatcatttcaaaaaatgtgtatttc
acatttcttggacttttttagctgttattcagtgatgcattttgtatac
tcacgtggtatttagtaataaaaatctatctatgtattacgtcac
(SEQ ID NO:4)

chr1_70_2399.c protein

SVVMVVFDNEKVPVEQLRFWKHWHSRQPTAKQRVIDVADCKENFNTVEHI
EEVAYNALSFVWNVNEEAKVFIGVNCLSTDFSSQKGVKGVPLNLQIDTYD
CGLGTERLVHRAVCQIKIFCDKGAERKMRDDERKQFRRKVKCPDSSNSGV
KGCLLSGFRGNETTYLRPETDLETPPVLFIPNVHFSSLQRSGGAAPSAGP
SSSNRLPLKRTCSPFTEEFEPLPSKQAKEGDLQRVLLYVRRETEEVFDAL
MLKTPDLKGLRNAISEKYGFPEENIYKVYKKCKRGILVNMDNNIIQHYSN
HVAFLLDMGELDGKIQIILKEL(SEQ ID NO:5)

chr1_70_2399.f mRNA sequence (coding in CAPITALS, no ATG at start)

aagttgccccacctctctgagcattggcttccccatctgtgaaagaggag
tgctgatgtttgccttctaggggcctagtgaggcttaagggtgagcagca
ggcacacagaaagctagaaatacaggatcactgtgggacggtggggctgg
ccacctgggcaggccacttacccagcggccccctctgtctccaggtgttc
atcggcgtaaactgtctgagcacagacttttcctcacaaaaggggtgaa
gggtgtccccctgaacctgcagattgacacctatgactgtggcttgggca
ctgagcgcctggtacaccgtgctgtctgccagatcaagatcttctgtgac
aagggagctgagaggaagatgcgcgatgacgagcggaagcagttccggag
gaaggtcaagtgccctgactccagcaacagtggcgtcaagggctgcctgc
tgtcgggcttcaggggcaatgagacgacctaccttcggccagagactgac
ctggagacgccacccgtgctgttcatccccaatgtgcacttctccagcct
gcagcggtctggaggggcagccccctcggcaggacccagcagctccaaca
ggctgcctctgaagcgtacctgctcgcccttcactgaggagtttgagcct
ctgccctccaagcaggccaaggaaggcgaccttcagagagttctgctgta
tgtgcggagggagactgaggaggtgtttgacgcgctcatgttgaagaccc
cagacctgaaggggctgaggaatgcgatctctgagaagtatgggttccct
gaaGAGAACATTTACAAAGTCTACAAGAAATGCAAGCGAGGAATCTTAGT
CAACATGGACAACAACATCATTCAGCATTACAGCAACCACGTCGCCTTCC
TGCTGGACATGGGGGAGCTGGACGGCAAAATTCAGATCATCCTTAAGGAG
CTGTAAggcctctcgagcatccaaaccctcacgacctgcaaggggccagc
agggacgtggccccacgccacacacaacctctccacatgcctcagcgctg
ttacttgaatgccttccctgagggaagaggcccttgagtcacagacccac
agacgtcagggccagggagagacctaggggtcccctggcctggatcccc
atggtatgcttgaatctgctccctgaacttcctgccagtgcctccccgta
ccccaaaacaatgtcaccatggttaccacctacccagaagactgttccct
cctcccaagacccttgtctgcagtggtgctcctgcaggctgcccgttaag
atggtggcggcacacgctccctcccgcagcaccacgccagctggtgcggc
ccccactctctgtcttccttcaacttcagacaaaggatttctcaaccttt
ggtcagttaacttgaaaactcttgattttcagtgcaaatgacttttaaaa
gacactatattggagtctctttctcagacttcctcagcgcaggatgtaaa
tagcactaacgatcgactggaacaaagtgaccgctgtgtaaaactactgc
cttgccactcactgttgtatacatttcttatttacgattttcatttgtta
tatatatatataaatatactgtatatatatgcaacattttatatttttca
tggatatgtttttatcatttcaaaaaatgtgtatttcacatttcttggac
ttttttttagctgttattcagtgatgcattttgtatactcacgtggtattt

agtaataaaatctatctatgtattacgtcac(SEQ ID NO:6)

chr1_70_2399.f protein

MRDDERKQFRRKVKCPDSSNSGVKGCLLSGFRGNETTYLRPETDLETPPV
LFIPNVHFSSLQRSGGAAPSAGPSSSNRLPLKRTCSPFTEEFEPLPSKQA
KEGDLQRVLLYVRRETEEVFDALMLKTPDLKGLRNAISEKYGFPEENIYK
VYKKCKRGILVNMDNNIIQHYSNHVAFLLDMGELDGKIQIILKEL
(SEQ ID NO:7)

C1000572 mRNA sequence (all coding, UTRs not shown)

ATGAAAAGGTCTGTGCGGCTGCTAAAGAACGACCCAGTCAACTTGCAGAA
ATTCTCTTACACTAGTGAGGATGAGGCCTGGAAGACGTACCTAGAAAACC
CGTTGACAGCTGCCACAAAGGCCATGATGAGAGTCAATGGAGATGATGAG
AGTGTTGCGGCCTTGAGCTTCCTCTATGATTACTACATGTCGATGCTCTT
CCCAGATATCCTGAAAACCTCCCCGGAACCCCCATGTCCAGAGGACTACC
CCAGCCTCAAAAGTGACTTTGAATACACCCTGGGCTCCCCCAAAGCCATC
CACATCAAGTCAGGCGAGTCACCCATGGCCTACCTCAACAAAGGCCAGTT
CTACCCCGTCACCCTGCGGACCCCAGCAGGTGGCAAAGGCCTTGCCTTGT
CCTCCAACAAAGTCAAGAGTGTGGTGATGGTTGTCTTCGACAATGAGAAG
GTCCCAGTAGAGCAGCTGCGCTTCTGGAAGCACTGGCATTCCCGGCAACC
CACTGCCAAGCAGCGGGTCATTGACGTGGCTGACTGCAAAGAAAACTTCA
ACACTGTGGAGCACATTGAGGAGGTGGCCTATAATGCACTGTCCTTTGTG
TGGAACGTGAATGAAGAGGCCAAGGTGTTCATCGGCGTAAACTGTCTGAG
CACAGACTTTTCCTCACAAAAGGGGGTGAAGGGTGTCCCCCTGAACCTGC
AGATTGACACCTATGACTGTGGCTTGGGCACTGAGCGCCTGGTACACCGT
GCTGTCTGCCAGATCAAGATCTTCTGTGACAAGGGAGCTGAGAGGAAGAT
GCGCGATGACGAGCGGAAGCAGTTCCGGAGGAAGGTCAAGTGCCCTGACT
CCAGCAACAGTGGCGTCAAGGGCTGCCTGCTGTCGGGCTTCAGGGGCAAT
GAGACGACCTACCTTCGGCCAGAGACTGACCTGGAGACGCCACCCGTGCT
GTTCATCCCCAATGTGCACTTCTCCAGCCTGCAGCGGTCTGGAGGGAGCC
TCCAGCAGCCAGGGGCTCCTCTCATTTTCCTGCGTGTGATGGAAAATGTC
TTTTTCACTTCATTGCAGGCAGCCCCCTCGGCAGGACCCAGCAGCTCCAA
CAGGCTGCCTCTGAAGCGTACCTGCTCGCCCTTCACTGAGGAGTTTGAGC
CTCTGCCCTCCAAGCAGGCCAAGGAAGGCGACCTTCAGAGAGTTCTGCTG
TATGTGCGGAGGGAGACTGAGGAGGTGTTTGACGCGCTCATGTTGAAGAC

CCCAGACCTGAAGGGGCTGAGGAATGCGATCTCTGAGAAGTATGGGTTCC
CTGAAGAGAACATTTACAAAGTCTACAAGAAATGCAAGCGAGGAATCTTA
GTCAACATGGACAACAACATCATTCAGCATTACAGCAACCACGTCGCCTT
CCTGCTGGACATGGGGGAGCTGGACGGCAAAATTCAGATCATCCTTAAGG
AGCTGTAA(SEQ ID NO:8)

C1000572 Protein

MKRSVRLLKNDPVNLQKFSYTSEDEAWKTYLENPLTAATKAMMRVNGDDE
SVAALSFLYDYYMSMLFPDILKTSPEPPCPEDYPSLKSDFEYTLGSPKAI
HIKSGESPMAYLNKGQFYPVTLRTPAGGKGLALSSNKVKSVVMVVFDNEK
VPVEQLRFWKHWHSRQPTAKQRVIDVADCKENFNTVEHIEEVAYNALSFV
WNVNEEAKVFIGVNCLSTDFSSQKGVKGVPLNLQIDTYDCGLGTERLVHR
AVCQIKIFCDKGAERKMRDDERKQFRRKVKCPDSSNSGVKGCLLSGFRGN
ETTYLRPETDLETPPVLFIPNVHFSSLQRSGGSLQQPGAPLIFLRVMENV
FFTSLQAAPSAGPSSSNRLPLKRTCSPFTEEFEPLPSKQAKEGDLQRVLL
YVRRETEEVFDALMLKTPDLKGLRNAISEKYGFPEENIYKVYKKCKRGIL
VNMDNNIIQHYSNHVAFLLDMGELDGKIQIILKEL(SEQ ID NO:9)

ctgChr_1ctg20.176 mRNA sequence (all coding, UTRs not shown)

ATGGAGGCAGGGGAGAAAAGCGCTCTGGGTGCCTGGAGCCCGCAGCCCTG
GGCAGCCCCGGGCTACCGCAGGGCGCAAGGGATCCTGGGCTGCGGCCGAG
GGCGCCGGAAGTCGCCGCCGACCGCCTGGGTCTCGCAGGAAAACAGCCGG
CGCCCGCGAGCTGCCCAGCGTCGGGTTTTCCTGAAGAGCCCAGCTCCTCA
CACCTTGGGGCCTGGTGGGATGGGAGACACTGTCCTGGATGAAGCCGCTG
GGAGAGCTGCCGCCTCCTGTATGCTGAGGTCTGTGCGGCTGCTAAAGAAC
GACCCAGTCAACTTGCAGAAATTCTCTTACACTAGTGAGGATGAGGCCTG
GAAGACGTACCTAGAAAACCCGTTGACAGCTGCCACAAAGGCCATGATGA
GAGTCAATGGAGATGATGAGAGTGTTGCGGCCTTGAGCTTCCTCTATGAT
TACTACATGGGTCCCAAGGAGAAGCGGATATTGTCCTCCAGCACTGGGGG
CAGGAATGACCAAGGAAAGAGGTACTACCATGGCATGGAATATGAGACGG
ACCTCACTCCCCTTGAAAGCCCCACACACCTCATGAAATTCCTGACAGAG
AACGTGTCTGGAACCCCAGAGTACCCAGATTTGCTCAAGAAGAATAACCT
GATGAGCTTGGAGGGGGCCTTGCCCACCCCTGGCAAGGCAGCTCCCCTCC
CTGCAGGCCCCAGCAAGCTGGAGGCCGGCTCTGTGGACAGCTACCTGTTA
CCCACCACTGATATGTATGATAATGGCTCCCTCAACTCCTTGTTTGAGAG

CATTCATGGGGTGCCGCCCACACAGCGCTGGCAGCCAGACAGCACCTTCA
AAGATGACCCACAGGAGTCGATGCTCTTCCCAGATATCCTGAAAACCTCC
CCGGAACCCCCATGTCCAGAGGACTACCCCAGCCTCAAAAGTGACTTTGA
ATACACCCTGGGCTCCCCCAAAGCCATCCACATCAAGTCAGGCGAGTCAC
CCATGGCCTACCTCAACAAAGGCCAGTTCTACCCCGTCACCCTGCGGACC
CCAGCAGGTGGCAAAGGCCTTGCCTTGTCCTCCAACAAAGTCAAGAGTGT
GGTGATGGTTGTCTTCGACAATGAGAAGGTCCCAGTAGAGCAGCTGCGCT
TCTGGAAGCACTGGCATTCCCGGCAACCCACTGCCAAGCAGCGGGTCATT
GACGTGGCTGACTGCAAAGAAAACTTCAACACTGTGGAGCACATTGAGGA
GGTGGCCTATAATGCACTGTCCTTTGTGTGGAACGTGAATGAAGAGGCCA
AGGTGTTCATCGGCGTAAACTGTCTGAGCACAGACTTTTCCTCACAAAAG
GGGGTGAAGGGTGTCCCCCTGAACCTGCAGATTGACACCTATGACTGTGG
CTTGGGCACTGAGCGCCTGGTACACCGTGCTGTCTGCCAGATCAAGATCT
TCTGTGACAAGGGAGCTGAGAGGAAGATGCGCGATGACGAGCGGAAGCAG
TTCCGGAGGAAGGTCAAGTGCCCTGACTCCAGCAACAGTGGCGTCAAGGG
CTGCCTGCTGTCGGGCTTCAGGGGCAATGAGACGACCTACCTTCGGCCAG
AGACTGACCTGGAGACGCCACCCGTGCTGTTCATCCCCAATGTGCACTTC
TCCAGCCTGCAGCGGTCTGGAGGGCTCCAACTGCCTAGTTACCGGCCGCA
GGACCATCTGCAATTCCCAGCCCTTCTGGGCATGCTGGGGCCCAGGCTGC
CTCTGAAGCGTACCTGCTCGCCCTTCACTGAGGAGTTTGAGCCTCTGCCC
TCCAAGCAGGCCAAGGAAGGCGACCTTCAGAGAGTTCTGCTGTATGTGCG
GAGGGAGACTGAGGAGGTGTTTGACGCGCTCATGTTGAAGACCCCAGACC
TGAAGGGGCTGAGGAATGCGATCTCTGAGAAGTATGGGTTCCCTGAAGAG
AACATTTACAAAGTCTACAAGAAATGCAAGCGAGGAATCTTAGTCAACAT
GGACAACAACATCATTCAGCATTACAGCAACCACGTCGCCTTCCTGCTGG
ACATGGGGGAGCTGGACGGCAAAATTCAGATCATCCTTAAGGAGCTGTAA
(SEQ ID NO:10)

ctgChr_1ctg20.176 protein

MEAGEKSALGAWSPQPWAAPGYRRAQGILGCGRGRRKSPPTAWVSQENSR
RPRAAQRRVFLKSPAPHTLGPGGMGDTVLDEAAGRAAASCMLRSVRLLKN
DPVNLQKFSYTSEDEAWKTYLENPLTAATKAMMRVNGDDESVAALSFLYD
YYMGPKEKRILSSSTGGRNDQGKRYYHGMEYETDLTPLESPTHLMKFLTE
NVSGTPEYPDLLKKNNLMSLEGALPTPGKAAPLPAGPSKLEAGSVDSYLL
PTTDMYDNGSLNSLFESIHGVPPTQRWQPDSTFKDDPQESMLFPDILKTS

PEPPCPEDYPSLKSDFEYTLGSPKAIHIKSGESPMAYLNKGQFYPVTLRT
PAGGKGLALSSNKVKSVVMVVFDNEKVPVEQLRFWKHWHSRQPTAKQRVI
DVADCKENFNTVEHIEEVAYNALSFVWNVNEEAKVFIGVNCLSTDFSSQK
GVKGVPLNLQIDTYDCGLGTERLVHRAVCQIKIFCDKGAERKMRDDERKQ
FRRKVKCPDSSNSGVKGCLLSGFRGNETTYLRPETDLETPPVLFIPNVHF
SSLQRSGGLQLPSYRPQDHLQFPALLGMLGPRLPLKRTCSPFTEEFEPLP
SKQAKEGDLQRVLLYVRRETEEVFDALMLKTPDLKGLRNAISEKYGFPEE
(SEQ ID NO:11)

CICO3 (bs432ms434-222)

The 222 bases of the +3 PCR sequence from Celera overlapped with the 3'UTR of two different hypothetical proteins in the BLAT database.

bs432ms434-222
GATCTGCAATCAGAACTATTGAACTTCTCCATTCAGACCGCCACTCACACCTATGGGAAAAG
GGTAATGTATCATCGGCTTAGCAACAGGGAATACTATTCGTATGATGGAAAATGGGGACAAA
AGGCTTTGGTACATAAAACATTATTCCTTCCTTGGCCTAAAAACTCATCGCCACCTACATTA
(SEQ ID NO:12)

chr19_53_399.c mRNA sequence
tctggagcagctgaaaaacaaggaagtgaaacagccaattcctgccttaa
ctaattaacccaccttacgacattccaccattatgacgtgttcctgccct
gccccaactgatcaatcgaccctgtgacattcttctggacaatgagtccc
atcatctctccaccatgcaccttgtgactccctcctctgctgacaacaga
taaccacctttaactgtaactttccacagcctaccccagccctataaagc
tgcccctctcctatctcccttcgctgactctcttttcagactcagcccac
ttgcacccaagtgaattaacagccttgttgctcacacaaagcctgtttag
gtggtcttctatacggacatgcttgacacttggtgccaaaatctgggcca
gggggactccttcgtgagaccggcccctgtcctggccctcattccgtga
agagatccacctgcgacctcgggtcctcagaccagcccaaggaacatctc
accaatttcaaatcggatctcctcggcttagtggctgaagactgatgctg
cccgatcgcctcagaagcccttggaccatcacagatgccgagcttcggg
taactcttacggtggaggattcccagccatatgaagacaccctagctgga
cgatcagtccttgtcaaaagtctgacccctcaaactctacagcctcaatg

gaccagaccctacccggtcatttatagcacaccaactgccgtccatctgc
aggaccctctccattgggttcaccattccagaataaagccatgcccatca
gacagccagcttgatctctcctcttcctcctggaagccacaagattaggc
cgagagccgatcagacaaacaacctacaacccttaagctcctggcagcgc
ccagccaaggccatgcttccttgcaacactccttccaaatggccatccca
gcatgcttccaagcaggcttcatccgttcctctggaccctcatctcttaa
gacctgccgcctataaaaaggattatatcttgagaccctatcctctaaaa
ttttttccacacccaaaacaaaaaatctctgggtcaaaagtctaaaacgc
ttaggctggcaaccatcagatccttgcccatggtgtcctcaagcctactc
tcatgaaatggacaacagtacacgcatatggggccagttccacatatttg
gcaaccagaccagcatccaggacaacacaaagatctgcaatcagaactat
tgaacttctccattcagaccgccactcacacctatgggaaaagggtaatg
tatcatcggcttagcaacagggaatactattcgtatgatggaaaatgggg
acaaaaggctttggtacataaaacattattccttccttggcctaaaaact
catcgccacctacattaaagctaatatgcctgattactgttttagagaa
cttattttattagggcagttccaagctcaaaaatacgctaactggcacct
tgttagctacataaaaatgcaccctagacccgaaacttactagactcatt
ataaaattttctttaaggtgtccacgcagtccctggtcacacttgaagca
gtccggagaaatatcagccctaccccagtaatccccagaaggaacttaca
ctttttttttaatcttttcctacaacttcatattttataaataaaaagaca
aaaatgtcaggcctgtgagctgaagcttagccattgtaacccctgtgacc
tgcacatatccgtccaggtggcctgcaggagccaagaagtctggagcagc
cgaaaaaccacaaagaagtgaaacagccagttcctgccttaactaattaa
cccaccttacgacattccaccattatgacttgtccaccattatgacttgt
tcctgccctgccccaactgatcaatcaaccctgtgacattcttctcctgg
acaatgagtcccatcatctctccaccatgcaccttgtgacccctcctct
gctgaggataaccacctttaactgtaactttccacgcctacccaagccct
ataaagctgcccctctcctatctcccttcactgactctcttttcggactc
agcccacttgcacccaagtgaattaacagccttgttgctcacacaaagcc
tgattgggtgtcttctatacggacacgcgtgacaggaacctcaacccaaa
ggcagtctgatgaggtgtctaagataaaagtagcggcacaaaggcttttg
taaacagaggcgtttcatgtggttttcctttcctttccttatatgtgaaa
aggtgacagaaaagaaatcttcctaaaagagtc (SEQ ID NO: 13)

chr19_53_399.c protein

MGPVPHIWQPDQHPGQHKDLQSELLNFSIQTATHTYGKRVMYHRLSNREY
YSYDGKWGQKALVHKTLFLPWPKNSSPPTLKLICLITVFRELILLGQFQA
QKYANWHLVSYIKMHPRPETY(SEQ ID NO:14)

chr19_53_399:b mRNA sequence

tctggagcagctgaaaaacaaggaagtgaaacagccaattcctgccttaa
ctaattaacccaccttacgacattccaccattatgacgtgttcctgccct
gccccaactgatcaatcgaccctgtgacattcttctggacaatgagtccc
atcatctctccaccatgcaccttgtgactccctcctctgctgacaacaga
taaccacctttaactgtaactttccacagcctaccccagccctataaagc
tgcccctctcctatctcccttcgctgactctcttttcagactcagcccac
ttgcacccaagtgaattaacagccttgttgctcacacaaagcctgtttag
gtggtcttctatacggacatgcttgacacttggtgccaaaatctgggcca
gggggactccttcgtgagaccggcccccctgtcctggccctcattccgtga
agagatccacctgcgacctcgggtcctcagaccagcccaaggaacatctc
accaatttcaaatcggatctcctcggcttagtggctgaagactgatgctg
cccgatcgcctcagaagcccctggaccatcacagatgccgagcttcggg
taactcttacggtggaggattcccagccatatgaagacaccctagctgga
cgatcagtccttgtcaaaagtctgacccctcaaactctacagcctcaatg
gaccagaccctacccggtcatttatagcacaccaactgccgtccatctgc
aggaccctctccattgggttcaccattccagaataaagccatgcccatca
gacagccagcttgatctctcctcttcctcctggaagccacaagattaggc
cgagagccgatcagacaaacaacctacaacccttaagctcctggcagcgc
ccagccaaggccatgcttccttgcaacactccttccaaatggccatccca
gcatgcttccaagcaggcttcatccgttcctctggaccctcatctcttaa
gacctgccgcctataaaaaggattatatcttgagaccctatcctctaaaa
ttttttccacacccaaaacaaaaaatctctgggtcaaaagtctaaaacgc
ttaggctggcaaccatcagatccttgcccatggtgtcctcaagcctactc
tcatgaaatggacaacagtacacgcatatggggccagttccacatatttg
gcaaccagaccagcatccaggacaacacaaagtatgttgtttgttgttag
agggcttgggacatttcactctttgccagcctcagcttaatccaggagac
aaagattatttttccttattatctcttctgcataggatctgcaatcagaac
tattgaacttctccattcagaccgccactcacacctatgggaaaagggta
atgtatcatcggcttagcaacagggaatactattcgtatgatggaaaatg
gggacaaaaggctttggtacataaaacattattccttccttggcctaaaa

actcatcgccacctacattaaagctaatatgcctgattactgtttttaga
gaacttattttattagggcagttccaagctcaaaaatacgctaactggca
ccttgttagctacataaaaatgcaccctagacccgaaacttactagactc
attataaaatttttctttaaggtgtccacgcagtccctggtcacacttgaa
gcagtccggagaaatatcagccctaccccagtaatccccagaaggaactt
acacttttttttaatcttttcctacaacttcatattttataaataaaaag
acaaaaatgtcaggcctgtgagctgaagcttagccattgtaacccctgtg
acctgcacatatccgtccaggtggcctgcaggagccaagaagtctggagc
agccgaaaaaccacaaagaagtgaaacagccagttcctgccttaactaat
taacccaccttacgacattccaccattatgacttgtccaccattatgact
tgttcctgccctgccccaactgatcaatcaaccctgtgacattcttctcc
tggacaatgagtcccatcatctctccaccatgcaccttgtgacccccctcc
tctgctgaggataaccacctttaactgtaactttccacgcctacccaagc
cctataaagctgcccctctcctatctcccttcactgactctcttttcgga
ctcagcccacttgcacccaagtgaattaacagccttgttgctcacacaaa
gcctgattgggtgtcttctatacggacacgcgtgacaggaacctcaaccc
aaaggcagtctgatgaggtgtctaagataaaagtagcggcacaaaggctt
ttgtaaacagaggcgtttcatgtggttttcctttccttttccttatatgtg
aaaaggtgacagaaaagaaatcttcctaaaagagtc(SEQ ID NO:15)

chr19_53_399.b protein
CCPIASEAPWTITDAELRVTLTVEDSQPYEDTLAGRSVLVKSLTPQTLQP
QWTRPYPVIYSTPTAVHLQDPLHWVHHSRIKPCPSDSQLDLSSSSWKPQD
(SEQ ID NO:16)

EXAMPLE 2

Four DNA sequences was identified as being overexpressed in colon carcinoma using the Gene Logic Gene Express Oncology Datasuite. The sequences were identified in a datasuite search comparing gene expression in colon tumors with expression in normal tissues. These sequences represent genes and encode antigens which are to be targeted for the development of colon cancer therapeutics.

A. Sequence Information

The nucleotide sequences of each candidate are listed below. The first sequence listed

for each candidate was obtained directly from the public NCBI database (www.ncbi.nlm.nih.gov.), and corresponds to the Genbank accession number listed in the Gene Logic database. The additional sequence information provided was obtained by sequencing EST clones corresponding to each candidate.

Candidate 1: Genbank Accession #W91975

W91975/IMAGE clone 415310 3'mRNA sequence

GGCTTCTAAGGTACATTATGTTTTACTTTAATAAATAAAAATTAACTT
GAAGAAAAATGCAGNGCCCTATTTAATTGCTCTGCATGAAATGTACAG
AAACGGCAACCTCTGCGATTCTAAGCACTGTGAACGCCCCAGCCACAC
CGTGTCAACAAACCGTGTGGCACTTGGGAGAAGGCAGGGGTGATTTAC
GANTAGTCATGTTTCGCCTCCACCCGAGTCACTGCCAAGGAGTGGACA
GTGACACTGAATAAGCATNCGGNGCACCTCCTTCGGGAAGGGACTTGG
CTGACATGGTAGGCCTTCCCACTGGAGCCTGTACTTTGTCTTGCTGGG
CAGCACTCCANTCATGGGAAGGAACAATGANCAAGGCGTGGTGGTGGG
GGTGNGTAGGCCTGAGCGCCGTTTTCCATGGTGACCTTCACTGAGCAG
GCAGCAGGCACTGATGGGCAGTTGAGNCTGGNAGGAGTCAGGTCCTGG
TCNTGCCTC TGGTGTAACGCAGCANGCCATCAAAGGT(SEQ ID NO:17)

IMAGE clone 194681: T3 & T7 sequencing consensus

AGAATTCGGCACGAGNTTTTTTTTCTCTTAGATCTCCAGGTTCCCTTCCTTACCCCGGGA
AGCCTTTCTTCATCCCACCGTCCTGGGGCGTTNCACAGTGCTTAGAATCGCAGAGGTTGC
CGTTTCTGTACATTTCATGCAGAGCAATTAAATAGGGCACTGCATTTTTCTTCAAGTTAA
TTTTTATTTATTAAAGTAAAACATAATGTACCTTAGAAGCCAGACAGTCCTACAAGCTTA
TTATGTTGTACAGCGGCGTTCCGTCCCCCTCCCCAGCCCTCTCTTTCTAGAGGCAGCCAA
TTTCAGCTGTCTCTCTCTGCTTACCTACATATTTCCATGTTTCTTGGTTCATCACCTGGT
GGCACCTTCAGTCTGGAAACACCTGCCCTTCACTTTAGGGGAATTGGGCCCCTGTTCGTT
TGATAAGTTTTCCTACCATTTTCTGATTTGTTTTTTCTTTCTGGAAAATGTATTAGTCAG
ATGTAGGCTTTTCTGGATTAATCCTTCAACTTTCCTTTCTTTCTTTCCCTTCCTGCCTGT
CTCCCTGTTCTTTCTTACACTTTCTCAGGGAGATTCTTGACTGTATTTTCCAACTTTGTA
TCGACCATTTTACTTTTCCTGCCATATTTTCAATGTTTACTGATGTTTCTCTGCCCTTTC
AGTGCATCCTGGTTTTATTTCATGTTAGACTGAATCCATGTGAAATTGATAACAGGTTTT
CAGCCCACACACACACACACAAAAAAAAAAAAAAAAAAAAAAAA(SEQ ID NO:18)

Candidate 2: Genbank Accession # AI694242

AI694242/IMAGE clone 2327838 3'mRNA sequence

TTTTGTTGGCTGAGGCGGTATTTTCCTTTTATTGCTGTTATGAGATT
CAACATTTTTTCCAGAAATAACTTCTGAAAAGTGTGCCTAGATTTTG
AACACTTGTGATCCTAACATGTGGTGAGAAAGGCTTTTCAAAACACA
CACGTGTGGACAGAGGTCCACACACGGATACGTGTGCACACACGGGT
GCCTTGGGCGTGCGTCTTCCAAAAGGGGCGAGTACAGCTATCAACTT
GTGACTTCCAGGAGGCCTGGGTTTGCCTACGAAGGGGCCGTGTTCCC
AGTTGGCGTTCACACGTGGTGTACACACACAGGCACAGGCACCGTGT
CCCAAGGCCATCTCCCAAGGGCACCCGCAGACACTGGGCAGCCTTCT
CCGAAGCTGTCAGTGTCCTTCCTCGTGAGAGGATGATGAAGAGGATG
TGGTTTCCGCCGCCTCATCCACAGGCCGGCTG (SEQ ID NO: 19)

IMAGE clone 2327838: T3 & T7 sequencing consensus

NAAAANGGCGCCNGNCCCANNTAAAATNNACCCNCCTAAAGGGGAAAAACTNNGGCGGCC
GCCTTCGTTTTTTTTTTTTTTTTTTTGTGGTGGCTGAGGCGGTATTTTCCTTTTATTGCT
GTTAAGAGATTCAACATTTTTTCCAGAAATAACTTCTGAAAAGGGGGCCTNAGATTTTGA
ACACTTGGGATCCTAACAGGGGGTGAGAAAGGCTTTTCAAAACACACNACGGGTGGACAG
AGGTCCACACACGGNATACGGGGGCACACACGGGTGCCTTGGGCGTGCGTCTTCCAAAAG
GGGCGAGNTACAGCTATCAACTTGTGACTTCCAGGAGGCCTGGGTTTGCCTACGAAGGGG
CCGNTGTTCCCAGTTGGCGTTCACACGTGGTGTACACACACAGGCACAGGCACCNGTGTC
CCAANGGCCATCTNCCCAAGGGCACCCGCAGACACTGGGCAGCCTTCTCCGAAGCTGTCA
GTGTCCTTCCTCGTGAGAGGATGATGAAGAGGATGTGGTTTCCGCCGCCTCATCCACAGG
CCGGCTGCCCACGGAGCCTTAGACATCGAGGCCAGAGCGACAGAAGCCTGTGTGCTGACC
GGCCTGGTCTCCTTTGACGTCTCGAGCAGCTTGGCAGGGTGGGAAAAGTAGCCTGAGAGT
GATCCCCGGGCAGTGTCCGAGGCTCTGCCGTCCCCACCCCCACAGGCATCCAGGGGAGAG
AAACAACCTGCGCCTGCGAGGCCGTGCGGACCCCGCTCCACTCACCCCGCCTGGGGGGCC
AGAACCACCTCCCAGGGGCTTCCGCCAGTGCCGCAGTTGCTGACCCCAGGCAAACCTCGC
CGCCTCCTGCCCCGGCGGGCCTGGGATTTGCGAATGTGTGAAGGCATTAGCTGCCAGTTG
TAACTGGAACCCAGCCTAGAGGCCTCACTCCTCCAGCAGGAAGCCTTGTAATGCAGCGAA
TCTGAACCCGGCCCAGCGTCCAGAGACAGGAAGCATTAATAGGAGCGAATGTGAACACTG
TTCGCGCCCTGGCTGCGATTTATTGCCGATTGTGGGGAAAACATCAGTTGGTTGCAGAGT
TTCATTCATCTTTAGGGACAGGACCGGTGTGTCTGGGTGGCAGTTTAGAGAGCTGGGACA

GTCGGCATCACTCTGGGTGGCTCCTCTCAANCCCTGGTGCCTCGTGCCGAATTCTGGCCT
CGAGGCATTCTNAGGGGCTNTATNC(SEQ ID NO:20)

Candidate 3: Genbank Accession # AI680111

AI680111/IMAGE clone 2252029 3' mRNA sequence
TTTTTTTTTTTTGTGGATAAATATATTAGCAAATGAATATATTTCTTAACATAGTGCCT
GATTCAAGCGTCTGTCTGGTTCAAATATAAATACCCATGTGGGTACCTAGGTGCTAGTC
TCCCCACTAACTGAGGGAAAAAGGTTCCCAGGTGGGGTCCTCTGCCCACTTTGCCACCA
CATTCACATTCCAAATGGGATAATGCCTGAGGGGCCATGAGTGGTCAGGCTGCCCTGGG
GTGAATGTCACCCTGATGAGGCCCATCAGCTCTTGTCCACTCAGTGAGGCCAGACTTGT
GCTCTAATCCACT(SEQ ID NO:21)

IMAGE clone 2324560 T7 sequencing
CTNTGTANAAAGCTGGGTACGCGTAAGCTTGGGCCCCTCGAGGGATACTCTAGAGCGGC
CGCCCTTTTTTTTTTTTTTTTGTGGATAAATATATTAGCAAATAAATATATTTCTTAACA
TAGTGCCTGATTCAAGCGTCTGTCTGGTTCAGATATAAATACCCATGTGGGTACCTAGG
TGCTAGTCTCCCCACTAACTGAGGGAAAAAGGTTCCCAGGTGGGGTCCTCTGCCCACTT
TGCCACCACATTCACATTCCAAATGGGATAATGCCTGAGGGGCCAAGAGTGGTCAGGCT
GCCCTGGGGTGAATGTCACCCTGATGAGGCCCATCAGCTCTTGTCCACTCAGTGAGGCC
AGACTTGTGCTCTAATCCACTCTCCTGTGGGTCCCTGGCCTGTATGGCTTATACTGGGG
AGCTGGGCCTCTGGGCTGTCCAAACCCAAGGGTCACACTTTGCTTTTCCTTTGTTGTCC
CCATTTTCCATCCTTGCTCTAAGACAAAACTTTTCCCAGAGAAGAACTCTTTGTTGTCC
CCGCTCAGCTGTAATTCTGCCTTTTCTACCTTCATTCCATCCTTCCTCTGCCCAGATAA
AGTCCAGCAGAAATTCCTCCTTTCTACCTCTCTGGGACTCTGAGACAGGAAATCTTCAA
GGAGGAGTTTTTCCCTCCCCACTATTCTTATTCTCAACCCCCAGAAGAACCAANGGCTG
CTGTACCCCCCTCAGGGACAGAACTCCACACTATANGGGGGAAAGNTTCANGGGACCCC
TTCCTTTTANTGCTCANGGCTCCACCTATGCTACTGGNTCCTTTTGGCAAAAAAGGNAA
ATGANAGAGCCAGGGGTTGCCCCNTGATGTAACANCCNTTACTGGGGANGGGNCCAANG
NNGGTGNTCAAAGNNCCCCNAGGAGGGAGGNGANAAGGGGTCATGNGTTCTGCTNAANC
CNCTGGTTGGTATAAANTTGANGNTTGGGGTGANGGAAACCAAAAANGGNTGGAAAAAG
NAAAACACCTTTNNAAACCCTGGGTACCNNANATAAGNTTTTGGCCCNAAAAANTCNGC
CNNCAAGGGATCCGCCCCNCCCCCCCAGGGAAAAANTTGGTTCCTNGGGNGAAAAGGAN
TTTNCCCCCCNCAAATTTTNNCCNAAAAGNTTTGGAANTTGNAAAANAAAAGGANCCTT
CCCCCCCCCNCCACAAAAAAAAAAAAAAAAAA(SEQ ID NO:22)

IMAGE Clone 2324560 SP6 sequencing

CNNTTNCAAAAAGCAGGCTGGTACCGGTCCGGAATTCCCGGGATATCGTCGACCCACGC
CGTCCGGTTTGCTGGTGTTGCTGAAATAACTCCAGCAGAAGGAAAATTAATGCAGTCCC
ACCCGCTGTACCTGTGCAATGCCAGTGATGACGACAATCTGGAGCCTGGATTCATCAGC
ATCGTCAAGCTGGAGAGTCCTCGACGGGCCCCCCGCCCCTGCCTGTCACTGGCTAGCAA
GGCTCGGATGGCGGGTGAGCGAGGAGCCAGTGCTGTCCTCTTTGACATCACTGAGGATC
GAGCTGCTGCTGAGCAGCTGCAGCAGCCGCTGGGGCTGACCTGGCCAGTGGTGTTGATC
TGGGGTAATGACGCTGAGAAGCTGATGGAGTTTGTGTACAAGAACCAAAAGGCCCATGT
GAGGATTGAGCTGAAGGAGCCCCCGGCCTGGCCAGATTATGATGTGTGGATCCTAATGA
CAGTGGTGGGCACCATCTTTGTGATCATCCTGGCTTCGGTGCTGCGCATCCGGTGCCGC
CCCCGCCACAGCAGGCCGGATCCGCTTCAGCAGAGAACAGCCTGGGCCATCAGCCAGCT
GGCCACCAGGAGGTACCAGGCCAGCTGCAGGCAGGCCCGGGGTGAGTGGCCAGACTCAG
GGAGCAGCTGCAGCTCAGCCCCTGTGTGTGCCATCTGTCTGGAGGGAGTTCTCTGAGGG
GGCAGGAGCTACGGGTCATTTCCCTGCCTCCATGAGTTCCATCGTAACTGTGTGGACCC
CTGGNTACATCAGCATCCGGACTTGCCCCCTCTTGCATGGTTCAACATCACANAGGGGA
GATCCNTTTTCCCNGTCCCTGGGAACCTCTNCNATCTTACCAAGAACCAGGGTCGGAAG
ACTCCCCCCCTCATTTCNCCAGCATCCCCGGCATGNCCCACTACACCNTCCCTGGTNGCC
TACCTGTTNGGGCCCTTCCCCGGAATGCAGGGGNTNGGGCCCCCNCNAACTGGGTCCTT
TCCTGCCNTCCAGGNAGCCAGGCATGGGCCCCCCGAATCACCCCTTCCCCNAANATGGA
NNATCCCCCGGGTTCCAGGAAAACAAACAACCNCTGGAAGGAANCCNNNACCCCNTNNC
CCNAAGGCTGGGGAANGNAACNCCCCCNATTCCCCNTNNANGANCCCTNNGTTTNCNCN
AGGCCCCTNACCCGGGCCNNGCCCCCNAAACAAAGGGANTTGANAAANT
SEQ ID NO:24)

These sequences correspond to hypothetical gene FLJ20315/Genbank Accession AK000322

AK000322

AAAAAAAAAAAACTTTAGAGAAAGGAAGGGCCAAAACTACGACTTGGCTTTCTGAAACG
GAAGCATAAATGTTCTTTTCCTCCATTTGTCTGGATCTGAGAACCTGCATTTGGTATTA
GCTAGTGGAAGCAGTATGTATGGTTGAAGTGCATTGCTGCAGCTGGTAGCATGAGTGGT
GGCCACCAGCTGCAGCTGGCTGCCCTCTGGCCCTGGCTGCTGATGGCTACCCTGCAGGC
AGGCTTTGGACGCACAGGACTGGTACTGGCAGCAGCGGTGGAGTCTGAAAGATCAGCAG
AACAGAAAGCTGTTATCAGAGTGATCCCCTTGAAAATGGACCCCACAGGAAAACTGAAT

CTCACTTTGGAAGGTGTGTTTGCTGGTGTTGCTGAAATAACTCCAGCAGAAGGAAAATT
AATGCAGTCCCACCCACTGTACCTGTGCAATGCCAGTGATGACGACAATCTGGAGCCTG
GATTCATCAGCATCGTCAAGCTGGAGAGTCCTCGACGGGCCCCCCGCCCCTGCCTGTCA
CTGGCTAGCAAGGCTCGGATGGCGGGTGAGCGAGGAGCCAGTGCTGTCCTCTTTGACAT
CACTGAGGATCGAGCTGCTGCTGAGCAGCTGCAGCAGCCGCTGGGGCTGACCTGGCCAG
TGGTGTTGATCTGGGGTAATGACGCTGAGAAGCTGATGGAGTTTGTGTACAAGAACCAA
AAGGCCCATGTGAGGATTGAGCTGAAGGAGCCCCCGGCCTGGCCAGATTATGATGTGTG
GATCCTAATGACAGTGGTGGGCACCATCTTTGTGATCATCCTGGCTTCGGTGCTGCGCA
TCCGGTGCCGCCCCCGCCACAGCAGGCCGGATCCGCTTCAGCAGAGAACAGCCTGGGCC
ATCAGCCAGCTGGCCACCAGGAGGTACCAGGCCAGCTGCAGGCAGGCCCGGGGTGAGTG
GCCAGACTCAGGGAGCAGCTGCAGCTCAGCCCCTGTGTGTGCCATCTGTCTGGAGGAGT
TCTCTGAGGGGCAGGAGCTACGGGTCATTTCCTGCCTCCATGAGTTCCATCGTAACTGT
GTGGACCCCTGGTTACATCAGCATCGGACTTGCCCCCTCTGCGTGTTCAACATCACAGA
GGGAGATTCATTTTCCCAGTCCCTGGGACCCTCTCGATCTTACCAAGAACCAGGTCGAA
GACTCCACCTCATTCGCCAGCATCCCGGCCATGCCCACTACCACCTCCCTGCTGCCTAC
CTGTTGGGCCCTTCCCGGAGTGCAGTGGCTCGGCCCCCACGACCTGGTCCCTTCCTGCC
ATCCCAGGAGCCAGGCATGGGCCCTCGGCATCACCGCTTCCCCAGAGCTGCACATCCCC
GGGCTCCAGGAGAGCAGCAGCGCCTGGCAGGAGCCCAGCACCCCTATGCACAAGGCTGG
GGAATGAGCCACCTCCAATCCACCTCACAGCACCCTGCTGCTTGCCCAGTGCCCCTACG
CCGGGCCAGGCCCCCTGACAGCAGTGGATCTGGAGAAAGCTATTGCACAGAACGCAGTG
GGTACCTGGCAGATGGGCCAGCCAGTGACTCCAGCTCAGGGCCCTGTCATGGCTCTTCC
AGTGACTCTGTGGTCAACTGCACGGACATCAGCCTACAGGGGGTCCATGGCAGCAGTTC
TACTTTCTGCAGCTCCCTAAGCAGTGACTTTGACCCCCTAGTGTACTGCAGCCCTAAAG
GGGATCCCCAGCGAGTGGACATGCAGCCTAGTGTGACCTCTCGGCCTCGTTCCTTGGAC
TCGGTGGTGCCCACAGGGGAAACCCAGGTTTCCAGCCATGTCCACTACCACCGCCACCG
GCACCACCACTACAAAAAGCGGTTCCAGTGGCATGGCAGGAAGCCTGGCCCAGAAACCG
GAGTCCCCCAGTCCAGGCCTCCTATTCCTCGGACACAGCCCCAGCCAGAGCCACCTTCT
CCTGATCAGCAAGTCACCGGATCCAACTCAGCAGCCCCTTCGGGGCGGCTCTCTAACCC
ACAGTGCCCCAGGGCCCTCCCTGAGCCAGCCCCTGGCCCAGTTGACGCCTCCAGCATCT
GCCCCAGTACCAGCAGTCTGTTCAACTTGCAAAAATCCAGCCTCTCTGCCCGACACCCA
CAGAGGAAAAGGCGGGGGGGTCCCTCCGAGCCCACCCCTGGCTCTCGGCCCCAGGATGC
AACTGTGCACCCAGCTTGCCAGATTTTTCCCCATTACACCCCCAGTGTGGCATATCCTT
GGTCCCCAGAGGCACACCCCTTGATCTGTGGACCTCCAGGCCTGGACAAGAGGCTGCTA
CCAGAAACCCCAGGCCCCTGTTACTCAAATTCACAGCCAGTGTGGTTGTGCCTGACTCC
TCGCCAGCCCCTGGAACCACATCCACCTGGGGAGGGGCCTTCTGAATGGAGTTCTGACA

CCGCAGAGGGCAGGCCATGCCCTTATCCGCACTGCCAGGTGCTGTCGGCCCAGCCTGGC
TCAGAGGAGGAACTCGAGGAGCTGTGTGAACAGGCTGTGTGAGATGTTCAGGCCTAGCT
CCAACCAAGAGTGTGCTCCAGATGTGTTTGGGCCCTACCTGGCACAGAGTCCTGCTCCT
GGGAAAGGAAAGGACCACAGCAAACACCATTCTTTTTGCCGTACTTCCTAGAAGCACTG
GAAGAGGACTGGTGATGGTGGAGGGTGAGAGGGTGCCGTTTCCTGCTCCAGCTCCAGAC
CTTGTCTGCAGAAAACATCTGCAGTGCAGCAAATCCATGTCCAGCCAGGCAACCAGCTG
CTGCCTGTGGCGTGTGTGGGCTGGATCCCTTGAAGGCTGAGTTTTTGAGGGCAGAAAGC
TAGCTATGGGTAGCCAGGTGTTACAAAGGTGCTGCTCCTTCTCCAACCCCTACTTGGTT
TCCCTCACCCCAAGCCTCATGTTCATACCAGCCAGTGGGTTCAGCAGAACGCATGACAC
CTTATCACCTCCCTCCTTGGGTGAGCTCTGAACACCAGCTTTGGCCCCTCCACAGTAAG
GCTGCTACATCAGGGGCAACCCTGGCTCTATCATTTTCCTTTTTTGCCAAAAGGACCAG
TAGCATAGGTGAGCCCTGAGCACTAAAAGGAGGGGTCCCTGAAGCTTTCCCACTATAGT
GTGGAGTTCTGTCCCTGAGGTGGGTACAGCAGCCTTGGTTCCTCTGGGGGTTGAGAATA
AGAATAGTGGGGAGGGAAAAACTCCTCCTTGAAGATTTCCTGTCTCAGAGTCCCAGAGA
GGTAGAAAGGAGGAATTTCTGCTGGACTTTATCTGGGCAGAGGAAGGATGGAATGAAGG
TAGAAAAGGCAGAATTACAGCTGAGCGGGGACAACAAAGAGTTCTTCTCTGGGAAAAGT
TTTGTCTTAGAGCAAGGATGGAAAATGGGGACAACAAAGGAAAAGCAAAGTGTGACCCT
TGGGTTTGGACAGCCCAGAGGCCCAGCTCCCCAGTATAAGCCATACAGGCCAGGGACCC
ACAGGAGAGTGGATTAGAGCACAAGTCTGGCCTCACTGAGTGGACAAGAGCTGATGGGC
CTCATCAGGGTGACATTCACCCCAGGGCAGCCTGACCACTCTTGGCCCCTCAGGCATTA
TCCCATTTGGAATGTGAATGTGGTGGCAAAGTGGGCAGAGGACCCCACCTGGGAACCT
TTTTCCCTCAGTTAGTGGGGAGACTAGCACCTAGGTACCCACATGGGTATTTATATCT
GAACCAGACAGACGCTTGAATCAGGCACTATGTTAAGAAATATATTTATTTGCTAATA
TATTTAT(SEQ ID NO:24)

The hypothetical protein encoded by this sequence is contained under Genbank Accession BAA91085, provided below:

BAA91085/hypothetical protein

MSGGHQLQLAALWPWLLMATLQAGFGRTGLVLAAAVESERSAEQKAVIRVIPLKMDPTG
KLNLTLEGVFAGVAEITPAEGKLMQSHPLYLCNASDDDNLEPGFISIVKLESPRRAPRP
CLSLASKARMAGERGASAVLFDITEDRAAAEQLQQPLGLTWPVVLIWGNDAEKLMEFVY
KNQKAHVRIELKEPPAWPDYDVWILMTVVGTIFVIILASVLRIRCRPRHSRPDPLQQRT
AWAISQLATRRYQASCRQARGEWPDSGSSCSSAPVCAICLEEFSEGQELRVISCLHEFH
RNCVDPWLHQHRTCPLCVFNITEGDSFSQSLGPSRSYQEPGRRLHLIRQHPGHAHYHLP

AAYLLGPSRSAVARPPRPGPFLPSQEPGMGPRHHRFPRAAHPRAPGEQQRLAGAQHPYA
QGWGMSHLQSTSQHPAACPVPLRRARPPDSSGSGESYCTERSGYLADGPASDSSSGPCH
GSSSDSVVNCTDISLQGVHGSSSTFCSSLSSDFDPLVYCSPKGDPQRVDMQPSVTSRPR
SLDSVVPTGETQVSSHVHYHRHRHHHYKKRFQWHGRKPGPETGVPQSRPPIPRTQPQPE
PPSPDQQVTGSNSAAPSGRLSNPQCPRALPEPAPGPVDASSICPSTSSLFNLQKSSLSA
RHPQRKRRGGPSEPTPGSRPQDATVHPACQIFPHYTPSVAYPWSPEAHPLICGPPGLDK
RLLPETPGPCYSNSQPVWLCLTPRQPLEPHPPGEGPSEWSSDTAEGRPCPYPHCQVLSA
QPGSEEELEELCEQAV(SEQ ID NO:25)

Candidate 4: Genbank Accession # AA813827

AA813827/IMAGE:1271704 3', mRNA sequence
TTTTTTTTTAAACATTAAGATTTTATTACAAACCAGGCATTATATATTTCTTTACACTT
AAGGAATAGATATGAAACAATCTTGGAGTAAAAATTAGAAGGCAACTTGCTTCAAGTTT
GTACCAAGTCAATCAAGCAGAAACCTGAAGAACCTTGTTTTAAGATGAGAGTCATTTAT
ACTTGGCAGGCATTTTCTTCCAATGAAAAAATAAAGTCAATGTGCCATTATCTTGACAC
TTATAAAAATGTTTATAAAAAGCATTTAGGCCATTGATTCTCACAGTTGGCTGAATATT
GGAATCACCTAGATTAAAAAAAATACTAATCCCTATACAACATCCCCAAAATTCAGATT
TAATTAGTGTAAGTTAGGCCCTGGGCATATAGGCTGTTTTAAAATTCCTCGGGTGAGTC
TAATGTGTA(SEQ ID NO:26)

IMAGE 1341074 T7 sequencing
CCCNNCNNCCNNNNNNGNNNNNCTTANCTCGCAGNCANAATTCGGCCACGCAGGGTCGC
CTTCGCCGCCATGGNACGCCACCGGGCGCTGACAGACCTATGGAGAGTCAGGGTGTGCC
TCCCGGGCCTTATCGGGCCACCAAGCTGTGGAATGAAGTTACCACATCTTTTCGAGCAG
GAATGCCTCTAAGAAAACACAGACAACACTTTAAAAAATATGGCAATTGTTTCACAGCA
GGAGAAGCAGTGGATTGGCTTTATGACCTATTAAGAAATAATAGCAATTTTGGTCCTGA
AGTTACAAGGCAACAGACTATCCAACTGTTGAGGAAATTTCTTAAGAATCATGTAATTG
AAGATATCAAAGGGAGGTGGGGATCAGAAAATGTTGATGATAACAACCAGCTCTTCAGA
TTTCCTGCAACTTCGCCACTTAAAACTCTACCACGAAGGTATCCAGAATTGAGAAAAAA
CAACATAGAGAACTTTTCCAAAGATAAAGATAGCATTTTTAAATTACGAAACTTATCTC
GTAGAACTCCTAAAAGGCATGGATTACATTTATCTCAGGAAAATGGCGAGAAAATAAAG
CATGAAATAATCAATGAAAGATCAAGAAAATGCAATTGATAATAGAGAACTAAGCCAGG
AAGATGTTGAAAGAAGNTTGGGAGATATGTTATTCTGATCCTACCTGCAAACCATTTTA
AGGTGTGCCCATCCCCTAGAAGNAAGTTCTTAAATCCCAAACCAGGTAATTCCCCCAAN
TANTTAATGNACAAACATGGNCCAATACAAGTTAANCCNGGGAGTAGTTNTTACTACAA

AACCAATTCNGATGACCTTCCCCCACNGGNTNTTTNNCTNGCCATGGAAANGNCCCTAC
CAAANTGGCCCAANAANNCANTGATTTGGAATAATCCNNCCTTTGGTTGGGATTNNANC
AAATTGANTCCNAANNATCCCCAAATANTTTNCNAAANNCTCCCTGANCCCNACCTANC
TTTGGAANTTNCCCAATTNTTTGGCAAACNTTTTGGGGANGGAAAGAATTCTCCGGATT
TNAGCCCTTNTGGCAAAGGNTNCACCTNNNTTNAATTTNAAGANNNACACCCTNGGNAA
ATNTAANGGGGCCCCCNNATTNTTTNAAATNCGCGGAANAAGNTCCCAGGNTCCCNTNT
TTCCCCCCAAAATNNNATTGGGATTCCTNACCCCCCCAN(SEQ ID NO:27)

IMAGE 1341074 T3 sequencing

CNNNNNANTGCGGCCGCTCATTTTTTTTTTTTTTTTTCTCTATGNAAGCAGACTGNAG
NAAGAAGGCACTCAGNTTGATTTGAAGGAATTCAAATTGTTTAAGTGAAGGAATTTTGA
AGACTGTGGATCATCTTGAATTTTATGTATCCCACTGGATCTATCTGAAACTGTGATGT
AGCCACAAACAACTACCAGGAAATGAAACAAAAATTAAGATGCAACTGTATGACAGTGG
ACAAAAATAAAACAAAAACAATAGTAAAGTTAAAAAATAAAGCATTACTATAGTATATA
TTGTTAGTATAGTATACACAGTAGTTGCTTAATTCAGAAGCCACTTAAATAGGACACAT
GCAACATTCGGTTACAAACGTGCAAGACAGATGAGTGGTTTTCCCATTTGTAATATAAC
TTTAAAAAATTATTTCAACAGCCTAATTAAATGGATTGAGCCAGAATACATTTAAAAAA
TCTGTTCTCAGTCTGCAAGTACTAGAAACCTCATAAATATAAGATAATTGTGGTATAAT
AAAATACATATATTTGATCTTTGTCCTTGGTACCTGGTATGGAGCTCCTAAAATCCTTG
AAATTTCCTGAATGATAGAAGTCTTTAGTTACTCATAACAAGCCTATTTCAGCGNTATC
CTGAGTTTCATGCCTAANGGTAACTGANGGCCNGGCCATGGGTTTGAATTTTCATCCAC
CAACTACAACCCTTGTGGGGAGGAGAAAGGGNCTAGAAATTNAAGTTCNNTTGGNCCAC
CAGTGACCCAATGAATTGGGTCCNGTCATGCCTTGGNTANTTAAACCTTCCAATTAAAA
CNCNTAAAACATGCNAGGCTGANGGGAGTTTTNTAGGGTNNNGGAANCCTTGNATGGGG
CTGGGNATCCCCGGATTGACCCAGAAANGGTAAAAAAAACNCTTNGGCCCCCCCCCCCC
CCCTNACCCGGGGNCTTGGGAAACCCCTCCCTTTGGCCNTTTNCTGGAGGNCNACCCTT
TTNAAATAAACTAAAAGCCATAGNTAAAGGGGCNTTTTNCTNNTTNCTGGGAANCTTGN
ANGGAATTTTTNGACCCNGGNAAGGGGNTTTGAGGGAAANCCCAANTNGGTAATTGGCN
GGGCGGGAATTTNNATACCCCCNGAACCCNATTNCNCGGAATTAAAAAAATTTNGGNNC
GGNCCCCTTTNTNTNNNCCAGGGGTNAAANTTCTCNAAANNANAAA (SEQ ID NO: 28)

IMAGE 1676529 T7 sequencing

AGCTCGNAGCCAGATTCGGCACGAGGGAGATTATATGTTTTATTTATCATTGTCTCTGC
ATATCTGGAACAACGAAAGGCACATAGCAGTTGCTAAATAAATATCTTTTGAATGAATA
TATGATTGCCTTATACTTCTTTTATATCCCCATCTTCTAATAGATTATGAAAACTAGAA

TTCAAAATATATATACTGAACAAATGAATGACTGAAGCAATTGGGGATAATATTTAAGG
CAAAACCAAATCTGATAAAATATACACATATTTTAAAAACACATACATATATATAAATA
GATCAAAAGTGGAAAAAGAATATATAAAAGAGTGCAACATTTGGCAGCTGAGAATTATT
TCATTGAGTTTTCAAATATTCTTCACATTCTTATACTTAGAAACAAAGAAGTAACCCCA
AACAACTAATTCATTAGCTAATATCTCAGAACTTGCACATTTGCAGATAAATTTTCTTT
TAAGAACAGAATTATAGTTTAATCCCTAACACAGCTCAGTTTTCAAAATTCAAGTAAAT
AAAATTTTAGCACACATCATGATAGCCTTACTGGNATAGCTGTGTTAAAAACAAAAAGT
ATTTGGTATCATCTATTGTTATGTGCTCTCAATTGAGATCTAGTTAGTTTCCTAAGAGT
CTCACATTGATANCTATTTTGGGCACTTCCTTACATAATGNGNTTATTTAGAAATACCT
TATTAATGACAGACTTCCTTTTGAGTAGCTACATTCTCAGATATGGCTNCATTTATCAA
AGTTCCCCNAGGATTACCTAATTTTAATTCCAGTTAGNTATCTAAACTACGGAACTTTN
GGNTTTCCTTAAANTCAACATTGGTTGCCTTGATTGGAAGGNTTGGCNCCCAAAAANGG
CGGNCNTCCCNCNCCCGGGGGTGGNAANTCTTTTCNTGAANNTNCCAAGGNNAATTCCC
TCCNGAAANCNGGNTTTAANTTTTTTNCCNTTTCCCCCTTNAANGGGAAACCCCCGGGT
TTTNAAAAAAATTTTTCCCAAAANATTCNNCCNATGGGCCCCTTTGGAAAGGNAAAAAN
TTTTTTGTCCCTTAAAAANCCCTGGNAACCNAATTTGGTTNANCAAATANAGGAAGG
(SEQ ID NO:29)

IMAGE 167529 T3 sequencing

GCGGCCGCTGGGCCTGNGTGTCGCCTTCGCCGCCATGGNCGCCACCGGGCGCTGACAGA
CCTATGGAGAGTCAGGGTGTGCCTCCCGGGCCTTATCGGGCCACCAAGCTGTGGAATGA
AGTTACCACATCTTTTCGAGCAGGAATGCCTCTAAGAAAACACAGACAACACTTTAAAA
AATATGGCAATTGTTTCACAGCAGGAGAAGCAGTGGATTGGCTTTATGACCTATTAAGA
AATAATAGCAATTTTGGTCCTGAAGTTACAAGGCAACAGACTATCCAACTGTTGAGGAA
ATTTCTTAAGAATCATGTAATTGAAGATATCAAAGGGAGGTGGGGATCAGAAAATGTTG
ATGATAACAACCAGCTCTTCAGATTTCCTGCAACTTCGCCACTTAAAACTCTACCACGA
AGGTATCCAGAATTGAGAAAAAACAACATAGAGAACTTTTCCAAAGATAAAGATAGCAT
TTTTAAATTACGAAACTTATCTCGTAGAACTCCTAAAAGGCATGGATTACATTTATCTC
AGGAAAATGGCGAGAAAATAAAGCATGAAATAATCAATGAAGATCAAGAAAATGCAATT
GATAATAGAGAACTAAGCCAGGAAGATGTTGAAGAAGTTTGGGAGATATGTTATTCTGA
TCTACCTGCAAACCATTTTAGGTGTGCCATCCCTAGAAGAAGTCATAAATCCCAAACAA
GTAATTCCCCAATATATAATGTACNACATGGCCAATACANGTAACGTGGGAGTAGTTAT
ACTACAAACAAATCAGATGACCTCCCTCACTGGGTATTATCTGCCATGAAGNGCCTAGC
AAATNGGCCAGAAGCATGATATGNAATAATCCACCTTTGNNGGATTTGACCGANATGTN
TTNGAACATCCCGATTATTTCTAAACCCCTGACCNCTNNTACTTTGAAATNANAATTAT

TGNAANCTTTGGGNTGCTNCNCCCTTTAAAGGGGTGCCNCCAAGCCTNNGTTNGTGNTG
TTACTNCCCCCAANCGAAAAGNNCNCTTTATGGGTGNTNCCCAAGAACAATNTNN
(SEQ ID NO:30)

These sequences correspond to hypothetical gene FLJ20354/Genbank Accession AK000361

AK000361

GTGCCGAGACTCACCACTGCCGCGGCCGCTGGGCCTGAGTGTCGCCTTCGCCGCCATGG
ACGCCACCGGGCGCTGACAGACCTATGGAGAGTCAGGGTGTGCCTCCCGGGCCTTATCG
GGCCACCAAGCTGTGGAATGAAGTTACCACATCTTTTCGAGCAGGAATGCCTCTAAGAA
AACACAGACAACACTTTAAAAAATATGGCAATTGTTTCACAGCAGGAGAAGCAGTGGAT
TGGCTTTATGACCTATTAAGAAATAATAGCAATTTTGGTCCTGAAGTTACAAGGCAACA
GACTATCCAACTGTTGAGGAAATTTCTTAAGAATCATGTAATTGAAGATATCAAAGGGA
GGTGGGGATCAGAAAATGTTGATGATAACAACCAGCTCTTCAGATTTCCTGCAACTTCG
CCACTTAAAACTCTACCACGAAGGTATCCAGAATTGAGAAAAAACAACATAGAGAACTT
TTCCAAAGATAAAGATAGCATTTTTAAATTACGAAACTTATCTCGTAGAACTCCTAAAA
GGCATGGATTACATTTATCTCAGGAAAATGGCGAGAAAATAAAGCATGAAATAATCAAT
GAAGATCAAGAAAATGCAATTGATAATAGAGAACTAAGCCAGGAAGATGTTGAAGAAGT
TTGGAGATATGTTATTCTGATCTACCTGCAAACCATTTTAGGTGTGCCATCCCTAGAAG
AAGTCATAAATCCAAAACAAGTAATTCCCCAATATATAATGTACAACATGGCCAATACA
AGTAAACGTGGAGTAGTTATACTACAAAACAAATCAGATGACCTCCCTCACTGGGTATT
ATCTGCCATGAAGTGCCTAGCAAATTGGCCAAGAAGCAATGATATGAATGATCCAACTT
ATGTTGGATTTGAACGAGATGTATTCAGAACAATCGCAGATTATTTTCTAGATCTCCCT
GAACGTCTACTTACTTTTGAATATTACGAATTATTTGTAAACATTTTGGTTGTTTGTGG
CTACATCACAGTTTCAGATAGATCCAGTGGGATACATAAAATTCAAGATGATCCACAGT
CTTCAAAATTCCTTCACTTAAACAATTTGAATTCCTTCAAATCAACTGAGTGCCTTCTT
CTCAGTCTGCTTCATAGAGAAAAAAACAAAGAAGAATCAGATTCTACTGAGAGACTACA
GATAAGCAATCCAGGATTTCAAGAAAGATGTGCTAAGAAAATGCAGCTAGTTAATTTAA
GAAACAGAAGAGTGAGTGCTAATGACATAATGGGAGGAAGTTGTCATAATTTAATAGGG
TTAAGTAATATGCATGATCTATCCTCTAACAGCAAACCAAGGTGCTGTTCTTTGGAAGG
AATTGTAGATGTGCCAGGGAATTCAAGTAAAGAGGCATCCAGTGTCTTTCATCAATCTT
TTCCGAACATAGAAGGACAAAATAATAAACTGTTTTTAGAGTCTAAGCCCAAACAGGAA
TTCCTGTTGAATCTTCATTCAGAGGAAAATATTCAAAAGCCATTCAGTGCTGGTTTTAA
GAGAACCTCTACTTTGACTGTTCAAGACCAAGAGGAGTTGTGTAATGGGAAATGCAAGT

CAAAACAGCTTTGTAGGTCTCAGAGTTTGCTTTTAAGAAGTAGTACAAGAAGGAATAGT
TATATCAATACACCAGTGGCTGAAATTATCATGAAACCAAATGTTGGACAAGGCAGCAC
AAGTGTGCAAACAGCTATGGAAAGTGAACTCGGAGAGTCTAGTGCCACAATCAATAAAA
GACTCTGCAAAAGTACAATAGAACTTTCAGAAAATTCTTTACTTCCAGCTTCTTCTATG
TTGACTGGCACACAAAGCTTGCTGCAACCTCATTTAGAGAGGGTTGCCATCGATGCTCT
ACAGTTATGTTGTTTGTTACTTCCCCCACCAAATCGTAGAAAGCTTCAACTTTTAATGC
GTATGATTTCCCGAATGAGTCAAAATGTTGATATGCCCAAACTTCATGATGCAATGGGT
ACGAGGTCACTGATGATACATACCTTTTCTCGATGTGTGTTATGCTGTGCTGAAGAAGT
GGATCTTGATGAGCTTCTTGCTGGAAGATTAGTTTCTTTCTTAATGGATCATCATCAGG
AAATTCTTCAAGTACCCTCTTACTTACTAGACTGCTAGTGGATAATAACATCTTGACTA
CTTAAAAAAGGGACATATTGAAAATCCTGGAGATGGACTATTTGCTCCTTTGCCTAACT
TACTCATACTGTAAGCAGATTAGTGCTCAGGAGTTTGATGAGCAAAAAGTTTCTACCTC
TCAAGCTGCAATTGCTAGAACTCTTTAGAAAATATTATTAAAATACAGGAGTTTACCTT
AAAGGAAAAAAAAAAAACAAAAAAAAAAAAAAAAA(SEQ ID NO:31)

The hypothetical protein encoded by this sequence is contained under Genbank Accession BAA91111, provided below:

BAA91111/Hypothetical protein

MESQGVPPGPYRATKLWNEVTTSFRAGMPLRKHRQHFKKYGNCFTAGEAVDWLYDLLRNNSN
FGPEVTRQQTIQLLRKFLKNHVIEDIKGRWGSENVDDNNQLFRFPATSPLKTLPRRYPELRK
NNIENFSKDKDSIFKLRNLSRRTPKRHGLHLSQENGEKIKHEIINEDQENAIDNRELSQEDV
EEVWRYVILIYLQTILGVPSLEEVINPKQVIPQYIMYNMANTSKRGVVILQNKSDDLPHWVL
SAMKCLANWPRSNDMNDPTYVGFERDVFRTIADYFLDLPEPLLTFEYYELFVNILVVCGYIT
VSDRSSGIHKIQDDPQSSKFLHLNNLNSFKSTECLLLSLLHREKNKEESDSTERLQISNPGF
QERCAKKMQLVNLRNRRVSANDIMGGSCHNLIGLSNMHDLSSNSKPRCCSLEGIVDVPGNSS
KEASSVFHQSFPNIEGQNNKLFLESKPKQEFLLNLHSEENIQKPFSAGFKRTSTLTVQDQEE
LCNGKCKSKQLCRSQSLLLRSSTRRNSYINTPVAEIIMKPNVGQGSTSVQTAMESELGESSA
TINKRLCKSTIELSENSLLPASSMLTGTQSLLQPHLERVAIDALQLCCLLLPPPNRRKLQLL
MRMISRMSQNVDMPKLHDAMGTRSLMIHTFSRCVLCCAEEVDLDELLAGRLVSFLMDHHQEI
LQVPSYLLDC (SEQ ID NO:32)

B. Gene Logic Electronic Northern Data

The 'electronic Northerns' depicting the gene expression profile of the above described

sequences as determined using the Gene Logic datasuite are shown Figures 1-4. The values along the y-axis represent expression intensities in Gene Logic units. Each blue circle on the figure represents an individual patient sample. The bar graph on the left of the figure depicts the percentage of each tissue type found to express the gene fragment. The total number of samples for each tissue type is as follows: colon tumor, tumor % above 50, 31; colon tumors, 45; normal breast, 37; normal colon, 30; normal esophagus, 18, normal kidney, 28; normal liver, 21; normal lung, 35; normal lymph node 10; normal ovary, 25; normal pancreas, 20; normal prostate, 20; normal rectum, 22; normal stomach, 25, 'colon tumor, tumor % above 50' refers to tumor samples for which at least 50% of each sample comprises malignant tissue, as determined by a pathologist. This sample set is a subset of 'colon tumors', which comprises all colon tumor samples contained within the Gene Logic database.

EXAMPLE 3

Genes which were identified to be overexpressed in colon cancer tissues were further analyzed. Specifically, the sequences and data analyzing from the 10 pairs of malignant and normal colon tissues described above, the following additional observations and predictions were made.

bs421ms433-258

As with the CICO genes, we identified the following sequences which are differentially expressed in colon cancer and ABI sequenced the 258 bases set forth below.

bs421ms433-258

GATCTCACTCAGCAGACAGCAGCAGCCCGGGAGCCTGAGCTCAGGAGGAACTCTTACCTGGA
AATTGGGAACTGTATGGAGACTCCAAACTGACTTCTTTCAAAAAACAAAAACAAAAAATTTT
TTTAGCTTTGACAAACACACAAAAGTGGTAATAAAGAGAGCCCTCCTTGTCAACCCAAAATG
TGAGCCCCCTGTGGCAAAACCACCCCCTACCCCATTA (SEQ ID NO.33)

These bases correspond to the 3'UTR and some of the final coding exon of the hypothetical protein bK175E3.C22.6 which is contained in the Celera Database, the sequence of which is set forth below.

>bK175E3.C22.6
cggccgcggggcccggcgcggcgcgggccaaggagacggcgttcgtggag
gtggtgctgttcgagtcgagcccaagcggcgattacaccacctacaccac
cggcctcacgggccgcttctcgcgggccggggccacgctcagcgccgagg
gcgagatcgtgcagatgcacccactgggcctatgtaataacaatgacgaa
gaggacttgtatgaatatggctgggtaggagtggtgaagctggaacagcc
agaattggacccgaaaccatgcctcactgtcctaggcaaggccaagcgag
cagtacagcggggagctactgcagtcatctttgatgtgtctgaaaaccca
gaagctattgatcagctgaaccagggctctgaagacccgctcaagaggcc
ggtggtgtatgtgaagggtgcagatgccattaagctgatgaacatcgtca
acaagcagaaagtggctcgagcaaggatccagcaccgccctcctcgacaa
cccactgaatactttgacatggggattttcctggctttcttcgtcgtggt
ctccttggtctgcctcatcctccttgtcaaaatcaagctgaagcagcgac

gcagtcagaattccatgaacaggctggctgtgcaggctctagagaagatg
gaaaccagaaagttcaactccaagagcaaggggcgccgggaggggagctg
tggggccctggacacactcagcagcagctccacgtccgactgtgccatct
gtctggagaagtacattgatggagaggagctgcgggtcatcccctgtact
caccggtttcacaggaagtgcgtggaccccctggctgctgcagcaccacac
ctgcccccactgtcggcacaacatcatagaacaaaagggaaacccaagcg
cggtgtgtgtggagaccagcaacctctcacgtggtcggcagcagagggtg
accctgccggtgcattaccccggccgcgtgcacaggaccaacgccatccc
agcctaccctacgaggacaagcatggactcccacggcaacccccgtcacct
tgctgaccatggaccggcacggggagcagagcctctattccccgcagacc
cccgcctacatccgcagctacccacccctccacctggaccacagcctggc
cgctcaccgctgcggcctggagcaccgggcctactcccccagcccaccctt
tccgcaggcccaagttgagtggccgcagcttctccaaggcagcttgcttc
tcccagtatgagaccatgtaccagcactactacttccagggcctcagcta
cccggagcaggaggggcagtccccacctagcctcgcacccccggggcccgg
cccgtgcctttcctccgagcggcagtggcagcctgctcttccccaccgtg
gtgcacgtggccccgccctcccacctggagagcggcagcacgtccagctt
cagctgctatcacggccaccgctcggtgtgcagtggctacctggccgact
gcccaggcagcgacagcagcagcagcagctccggccagtgccactgt
tcctccagtgactctgtggtagactgcactgaggtcagcaaccagggcgt
gtacgggagctgctccaccttccgcagctccctcagcagcgactatgacc
ccttcatctaccgcagccggagcccctgtcgtgccagtgaggcgggggc
tcgggcagctcgggccggggacctgccctgtgcttcgagggctccccgcc
tcccgaggagctcccggcggtgcacagtcatggtgctgggcggggcgagc
cttggccgggccctgcctctccctcgggggatcaggtgtccacctgcagc
ctggagatgaactacagcagcaactcctccctggagcacaggggggcccaa
tagctctacctcagaagtggggctcgaggcttctcctggggccgcccctg
acctcaggaggacctggaaggggggccacgagttgccgtcgtgtgcctgc
tgctgcgagccccagccctccccagccgggcctagcgccggagcagctgg
cagcagcaccttgttcctggggccccacctctacgagggctctggcccgg
cgggtggggagccccagtcaggaagctcccagggcttgtacggccttcac
cccgaccatttgcccaggacagatggggtgaaatacgagggtctgccctg
ctgcttctatgaagagaagcaggtggcccgcgggggcggaggggcagcg
gctgctacactgaggactactcggtgagtgtgcagtacacgctcaccgag
gaaccaccgcccggctgctaccccgggcccgggacctgagccagcgcat
ccccatcattccagaggatgtggactgtgatctgggcctgccctcggact
gccaagggacccacagcctcggctcctggggtgggacgcgaggcccggat
accccacggccccacaggggcctgggagcaacccgggaagaggagcgggc
tctgtgctgccaggctagggccctactgcggcctggctgccctccggagg
aggcgggtgctgtcagggccaacttccctagtgccctocaggacactcag
gagtccagcaccactgccactgaggctgcaggaccgagatctcactcagc
agacagcagcagcccgggagcctgagctcaggaggaactcttacctggaa
attgggaactgtatggagactccaaactgacttctttcaaaaaacaaaaa
caaaaaatttttttagctttgacaaacacacaaaagtggtaataaagaga
gccctccttgtcaacccaaaatgtgagccccctgtggcaaaaccaccccc
taccccattaacaaatcaacagacaaaattctccgagtcctttgcctctt
ttgataacatgttgttctgttttgtaaagtgtgtgtgcttggggttccga
ggtgtgggattgagttctctgctttgtttttttttaagatattgtatgta
aatgtaaaaagttatttaaatatatatttaaagaaccctaactgccaac
ttttgctgaaaaagaaaaaaaaatcactgctgcattaaatgaaccacatc
atgtgtagatactgttgtctccctgaagggagctcaggcctttgaaaagc
tcagggcttcacctgccttagaaaatgaaccagaaacttgaagtaaagct
agttgataggggtacaggctctgaggagcagtgcaaaactgcctctttct
ttctcgtggcaaatcccaatgtacacgatttcaggtctcagacgccatgc
ctctccagcccacgcctttaggcaggtgatggcagcagctaggaataggg
tgtacatgatccacagccctgcggagccaggtcaagccgctgctatgaaa
gctccagggtgatggggacgattctgcccagtgtcctcagtctgtcccct
caggtcatggtcccaagtgaaatgacagagttcacagccctggtcttggc
tgaggtccaggtcatagtaagggcatgttcttggggccctcgacctgaac
tctgaccctccgggcagggaagaggaggttgtcccctttggttgtcctgg
ctttggagtcctttgcaaaaatattttgggccccctgccactggctgcag
aaatggctcgacggggtgtgtggggacagacacccagaaggaatgtactt
ttgtggccttggtgtccgatggggctgggggagagtgctctccactgacc
cagcagcacacccatgtgcagtgcgcctgcatctgtgtgggggcagccac

acccottggctgctgcttccttgggctgcctttctgggggcatgtgactg
gacctacgaggtctgcactgagctccatttgaatgatacctttcctatcc
catttcccccacggaagcaccgcttcagggttattcagtcctctgcctca
tggctgaaattgctcatctcgtctgcagatgtctactatcctgtctacct
aatgcactattatgtattgattctccatgagacagagagagagagagact
atcagatagtttacacccaaagggtaggttttgtatatttttccagcct
tttttattaaggggaaggggagagtttaaaaacccaaaccgttgtggttt
taaggtgtttcatttttaaaagggagagagaatctatttaaagctatttc
agatcagggattgtcatcctttttgtccaatgtattccttgttctttaa
aaaaattttttttagaggaaactaatattagtctttgtgttcactaactc
ttctggtcacttgtatttatttattcattcattcatcagatatttgttgc
catctgaaagaactggcccagtgggtctgaaagctcgcttgagaatagga
aacttgagacctggccccctgtgggtaggagaacaaggaccacctgggtt
ctccagtcttgaacgagaatctcactcttatcagaatgtttttcttaacc
tcagcgtatgatgaggaaatttacttatctctagctaggatttgacaaat
tccaacatcaaatgatcaaaacatttgccactgaggcttcactggtgaga
tccgttctccgtcctcgggtgcagtcccttgggggctgctcctcggactg
cgccccgcacacctgttatcgagggtgtgagaagcgcctaagctggtgac
atgtgatctgggacgccttcatttctcgggccaggagtagcagctgctaa
ggacagcagcttgcattgcgtggttttagggaagcagggtctggctttta
atatgaactgcaaaaagcagcttctcactgatatttttttgttgttgttt
ctgggggtttttttgttttgtttttaatgcctttgagtgcatattttct
tcctcgtctgaaaccgaactcccaaagtggctttctttagccctggctgg
aaaaccacctctcaatagccttaagcaataaatagatgagtagagaatgt
ggcttcaactgggcttattaaagtaagtgtgtctagttttcacttgaaca
agtgatagctgcagatggcgaaagaaacccatttaatttttgtagcttac
aggtggtagaaacaaaaatgcaattttaaaccttaaataccaaatacca
accattgccttttttttttttgagatggaattttgctcttgtcacccagg
ctggagtgcaatggcgcgatctcacctcactgcaacctctgcctcccggg
tccaagtgattctcctgcctcagcctcccaagtagctgggattacaggca
tgcgccaccacacccagctaattttgtatttttggtagagacagggtatc
tccatgttggtcaggctggtcttggattcccgacctcaggtgatccgccc
acctcggcctcccaaagtgctgggattacaggcgtgagccaccatgcctg
cccagcaataccaaccattgtcttttaaattcgtgttggcttctcagaca
gggagatcactggaataaaataaccgatggtcttattttgtcacacgtaa
atcaaaagaaatgtcctctttgaagttgtaagactccaccaatgacagac
acccttttcggtggactctgagtggtgtgtagtggttttatagccatgga
aactaggagtatctcactttccactgagaacccctgcccccaatccctct
aagttggggtgtggcagttgggcagggtcaagtgacccagccctggctgt
aggacagccatatacagtgaagagttctagaaccagctaaaaatggaagt
ttgggtgtttaccaacaaggtacctctttatggatgcagccccagtaagc
tggctttaactctcagctccttccctgtctcctcctaatccaagcccttt
tataaaataaagccccttctgtcccactgctcacatacttatgtgctgct
agtctctactcgaagttcgtgcaggactaatgcttttaaaatgaggtcta
aaaaataattactagtcgagactattattctttaaacagaactgcctttt
tctactctttatgtaaactctttctattgtgttggtctaacaaggcacta
ttttaaaatttttttaatttttcccatagcacttaaaagagattttgtaaa
gaccttgctgtaaagatttgtaataaaatggtctaagggctctttttcc
aacattaccatttttaaaaaatgttttaaaagctagaagacaacttatgt
atattctgtatatgtatagcagcacatttcatttatggaaatatgttctc
agaatatttatttactaatatatttatcttaagccatgtcttatgttgag
agtgtgacattgttggaataatcattgaaaatgactaacacaagaccctg
taaatacatgataattgcacacagattttacatatttgcagaccaaaaat
gatttaaaacaagttgtagtcttctatggttttgtaacaaattgtacaca
tgactgtaaaaaaaaaatacaattttatcaagtatgtgttata (SEQ ID NO. 34)

The above sequence encodes the following protein:

bK175E3.C22.6
MHPLGLCNNNDEEDLYEYGWVGVVKLEQPELDPKPCLTVLGKAKRAVQRG
ATAVIFDVSENPEAIDQLNQGSEDPLKRPVVYVKGADAIKLMNIVNKQKV
ARARIQHRPPRQPTEYFDMGIFLAFFVVVSLVCLILLVKIKLKQRRSQNS

MNRLAVQALEKMETRKFNSKSKGRREGSCGALDTLSSSSTSDCAICLEKY
IDGEELRVIPCTHRFHRKCVDPWLLQHHTCPHCRHNIIEQKGNPSAVCVE
TSNLSRGRQQRVTLPVHYPGRVHRTNAIPAYPTRTSMDSHGNPVTLLTMD
RHGEQSLYSPQTPAYIRSYPPLHLDHSLAAHRCGLEHRAYSPAHPFRRPK
LSGRSFSKAACFSQYETMYQHYYFQGLSYPEQEGQSPPSLAPRGPARAFP
PSGSGSLLFPTVVHVAPPSHLESGSTSSFSCYHGHRSVCSGYLADCPGSD
SSSSSSSGQCHCSSSDSVVDCTEVSNQGVYGSCSTFRSSLSSDYDPFIYR
SRSPCRASEAGGSGSSGRGPALCFEGSPPPEELPAVHSHGAGRGEPWPGP
ASPSGDQVSTCSLEMNYSSNSSLEHRGPNSSTSEVGLEASPGAAPDLRRT
WKGGHELPSCACCCEPQPSPAGPSAGAAGSSTLFLGPHLYEGSGPAGGEP
QSGSSQGLYGLHPDHLPRTDGVKYEGLPCCFYEEKQVARGGGGGSGCYTE
DYSVSVQYTLTEEPPPGCYPGARDLSQRIPIIPEDVDCDLGLPSDCQGTH
SLGSWGGTRGPDTPRPHRGLGATREEERALCCQARALLRPGCPPEEAGAV
RANFPSALQDTQESSTTATEAAGPRSHSADSSSPGA (SEQ ID NO. 35)

This protein contains a transmembrane domain as determined by SMART (shown below), SOSUI, and TmPred. SMART also predicts that this protein contains a RING domain; these domains are zinc finger domains involved in protein: protein interactions. The structure of the protein is depicted schematically below:

EXAMPLE 4

Using the GeneLogics database and the methods described generally in example 2, we identified additional DNA sequences that are upregulated in colon tumor tissues which are identified below.

AA781143/Hs19_11415_28_1_1699a

We found fragment AA781143 was upregulated 4.16 fold in the colon samples compared to mixed normal tissue. As shown in Figure 6, Enorthern analysis of this fragment demonstrates that it is expressed in 69% of the colon tumors with greater than 50% malignant cells and is expressed very little in normal tissues.

DNA sequence for AA781143

TTGTCTTCTACGACCAGCTGAAGCAAGTGATGAATGCGTACAGAGTCAAGCCGG
CCGTCTTTGACCTGCTCCTGGCTGTTGGCATTGCTGCCTACCTCGGCATGGCCTAC
GTGGCTGTCCAGGTGAGCAGTGCCCAGGCTCAGCACTTCAGCCTCCTCTACAAGA
CCGTCCAGAGGCTGCTCGTGAAGGCCAAGACACAGTGACACAGCCACCCCCACA
GCCGGAGCCCCCGCCGCTCCACAGTCCCTGGGGCCGAGCACGAGTTGGNAGGGG
ACCCTCTTCTCCCGTCNTGCCNTCGGGTTGCCCGCCTCCTCCAGAGACTTNNCAA
GGGCCCATCACCACTGGCCTCTGGGCACTTGTGCTGAGACTCTGGGACCCAGGCA
GCTGCCACCTTGTCACCATGAGAGAATTTGGGGAGTGCTTGCATGCTAGCCAGCA

GGCTCCTGTCTGGGTGCCACGGGGCCAGCATTTTGGAGGGAGCTTCCTTCCTTCC
TTCCTGGACAGGTCGTCATGATGGATGCACTGACTGACCGTCTGGGGCTCAGGCT
GGTGTGGGATGCAGCCGGCCG (SEQ ID NO: 36)

The GeneLogic database calls this protein "hypothetical protein from EUROIMAGE 2021883."

Nucleotide sequence:

CCAGAGTTTGTCTTCTACGACCAGCTGAAGCAAGTGATGAATGCGTACAGAGTCA
AGCCGGCCGTCTTTGACCTGCTCCTGGCTGTTGGCATTGCTGCCTACCTCGGCAT
GGCCTACGTGGCTGTCCAGCACTTCAGCCTCCTCTACAAGACCGTCCAGAGGCTG
CTCGTGAAGGCCAAGACACAGTGACACAGCCACCCCCACAGCCGGAGCCCCCGC
CGCTCCACAGTCCCTGGGGCCGAGCACGAGTGAGTGGACACTGCCCCGCCGCGG
GCGGCCCTGCAGGGACAGGGGCCCTCTCCCTCCCCGGCGGTGGTTGGAACACTG
AATTACAGAGCTTTTTTCTGTTGCTCTCCGAGACTGGGGGGGGATTGTTTCTTCTT
TTCCTTGTCTTTGAACTTCCTTGGAGGAGAGCTTGGGAGACGTCCCGGGGCCAGG
CTACGGACTTGCGGACGAGCCCCCCAGTCCTGGGAGCCGGCCGCCCTCGGTCTG
GTGTAAGCACACATGCACGATTAAAGAGGAGACGCCGGGACCCCCTGCCCGATC
GCGCGCGGCCTCCGCCCACCGCCTCCTGCCGCAAGGGGCCTGGACTGCAGGCCT
GACCTGCTCCCTGCTCCGTGTCTGTCCTAGGACGTCCCCTCCCGCTCCCCGATGGT
GGCGTGGACATGGTTATTTATCTCTGCTCCTTCTTGCCTGGAGGAGGGCAGTGCC
AGCCCTGGGGTTCTGGGATTCCAGCCCTCCTGGAGCCTTTTGTTCCCCATGTGGTC
TCAGTGACCCGTCCCCCTGACAGTGGGCTCGGGGAGCTGCATCACCCAGCCTTCC
CCTTCTCCGACTGCAGGGTCTGATGTCATCATTGACAGCCTTTGCTTCGTGGGGG
CCTGGCAGGGCCCCTGCCTCCCCGACCCCCGACCCACTGCAAATCCCCGTTCCCC
TGCACTCCTCTTCTCCCAGCCCATCCCTCCGGCCCCTGTGCCTCTGCGGCCCCAGC
CCAGCTCCCAGGGCCGTCACCTGCTTGGCCCTGGCCCAGCTCCCTGCCCTGAGTC
CTGAGCCAGTGCCTGGTGTTTCCTGGGCTCGGTACTGGGCCCCCAGGCCATCCAG
GCTTTGCCACGGCCAGTTGGTCCTCCCTGGGGAACTGGGTGCGGGTGGAGTACTG
GGAGGCAGGAGGTGGCCCGGGGAGGCCTTGTGGCTCCTCCCCTCGCTCCTCGCCC
TGGGCCTCAGCTTCCTCATCAATAGAAAGGATGTGTTCGGGGTGGGGGCGTCAG
GTGAGAACGTTTGCTGGGAAGGAGAGGACTTGGGGCATGGCCTCTGGGGCCACC
CTTCCTGGAACTCAGAGAGGAAGGTCCGGGCCCTCGGGAAGCCTTGGACAGAAC
CCTCCACCCCGCAGACCAGGCGTCGTGTGTGTGTGGGAGAGAAGGAGGCCCGTG
TTGAGCTCAGGGAGACCCCGGTGTGTCCGTTCTTAGCAATATAACCTACCCAGTG
CGTGCCGAGCAGGCTTGGTGGGGAAGGGACTTGAGCTGGGCAAGTCCTGGCCTG
GCACCCGCAGCCGTCTCCCTTCCGTGGCCCAGGGAGGTGTTTGCTGTCCGAAGGA
CCTGGGCCGGCCCATGGGAGCCTGGGGTTCTGTCCAGATAGGACCAGGGGGTCT
CACTTTGGCCACCAGTTCTTCGGCCAGCACCTCTGCCCTCCAGAACCTGCAGCCT
GGAGGGGTGAGGGGACAACCACCCCTCTTTCCTCCAGGTTGGCAGGGGACCCTC
TTCTCCCGTCTGCCCTGCGGGTTGCCCGCCTCCTCCAGAGACTTGCCCAAGGGCC
CATCACCACTGGCCTCTGGGCACTTGTGCTGAGACTCTGGGACCCAGGCAGCTGC
CACCTTGTCACCATGAGAGAATTTGGGGAGTGCTTGCATGCTAGCCAGCAGGCTC
CTGTCTGGGTGCCACGGGGCCAGCATTTTGGAGGGAGCTTCCTTCCTTCCTTCCT
GGACAGGTCGTCATGATGGATGCACTGACTGACCGTCTGGGGCTCAGGCTGGTGT
GGGA TGCAGCCGGCCGATGAGAAAATAAAGCCATATTGAATGAT
(SEQ ID NO: 37)

Protein Sequence

PEFVFYDQLKQVMNAYRVKPAVFDLLLAVGIAAYLGMAYVAVQHFSLLYKTVQRL LVKAKTQ (SEQ ID NO 38)

The protein set forth above contains one TM by SMART, SOSUI, and TmPred prediction programs. However, the BLAT database and EST sequences suggest that an alternative protein correspond thereto is Hs19_11415_28_1_1699a, the sequence of which as set forth below:

Nucleotide Sequence

>Hs19_11415_28_1_1699.a
gcaaggtcacgtcctgtccccacctttcgcccctcaccctagctcccca
acgccaaagacaaggttaagaaagtgatatcgcgaaatagttttttaaag
cattttattgcattttatgacttggagtttatgtgaaacctcaacggtat
tagccgaacagcctgccgcaccttccgggagttccagagtgggcctacaa
ctcccacagggctccgcgagcgccggacggacggactacaattcccgaca
ggcagcgcggctggcggggcggttcgccgcggtgcccacaggacctcagg
gcgagtgcgggctgccccgcgcggcgcccgcaggaccccggcggctaccc
atgccgaggtgagtccgcgggagccgccgccgccgccgtcccgtcccagc
tgccgccccgcgcggccccgccgccggccaggATGCTGGAGGAAGCGGGC
GAGGTGCTGGAGAACATGCTGAAGGCGTCTTGTCTGCCGCTCGGCTTCAT
CGTCTTCCTGCCCGCTGTGCTGCTGCTGGTGGCGCCGCCGCTGCCTGCCG
CCGACGCCGCGCACGAGTTCACCGTGTACCGCATGCAGCAGTACGACCTG
CAGGGCCAGCCCTACGGCACACGGAATGCAGTGCTGAACACGGAGGCGCG
CACGATGGCGGCGGAGGTGCTGAGCCGCCGCTGCGTGCTCATGCGGCTAC
TGGACTTCTCCTACGAGCAGTACCAGAAGGCCCTGCGGCAGTCGGCGGGC
GCCGTGGTCATCATCCTGCCCAGGGCCATGGCCGCCGTGCCCCAGGACGT
CGTCCGGCAATTCATGGAGATCGAGCCGGAGATGCTGGCCATGGAGACCG
CCGTCCCCGTGTACTTTGCCGTGGAGGACGAGGCCCTGCTGTCTATCTAC
AAGCAGACCCAGGCTGCCTCCGCCTCCCAGGGCTCCGCCTCTGCTGCTGA
AGTACTGCTGCGCACGGCCACTGCCAACGGCTTCCAGATGGTCACCAGCG
GGGTACAGAGCAAGGCCGTGAGTGACTGGCTGATTGCCAGCGTGGAGGGG
CGGCTGACGGGGCTGGGCGGAGAGGACCTTCCCACCATCGTCATCGTGGC
CCACTACGACGCCTTTGGAGTGGCCCCCTGGCTGTCGCTGGGCGCGGACT
CCAACGGGAGCGGCGTCTCTGTGCTGCTGGAGCTGGCACGCCTCTTCTCC
CGGCTCTACACCTACAAGCGCACGCACGCCGCCTACAACCTCCTGTTCTT
TGCGTCTGGAGGAGGCAAGTTTAACTACCAGGGAACCAAGCGCTGGCTGG
AAGACAACCTGGACCACACAGACTCCAGCCTGCTTCAGGACAATGTGGCC
TTCGTGCTGTGCCTGGACACCGTGGGCCGGGGCAGCAGCCTGCACCTGCA
CGTGTCCAAGCCGCCTCGGGAGGGCACCCTGCAGCACGCCTTCCTGCGGG
AGCTGGAGACGGTGGCCGCGCACCAGTTCCCTGAGGTACGGTTCTCCATG
GTGCACAAGCGGATCAACCTGGCGGAGGACGTGCTGGCCTGGGAGCACGA
GCGCTTCGCCATCCGCCGACTGCCCGCCTTCACGCTGTCCCACCTGGAGA
GCCACCGTGACGGCCAGCGCAGCAGCATCATGGACGTGCGGTCCCGGGTG
GATTCTAAGACCCTGACCCGTAACACGAGGATCATTGCAGAGGCCCTGAC
TCGAGTCATCTACAACCTGACAGAGAAGGGGACACCCCCAGACATGCCGG
TGTTCACAGAGCAGATGCAGATCCAGCAGGAGCAGCTGGACTCGGTGATG
GACTGGCTCACCAACCAGCCGCGGGCCGCGCAGCTGGTGGACAAGGACAG
CACCTTCCTCAGCACGCTGGAGCACCACCTGAGCCGCTACCTGAAGGACG
TGAAGCAGCACCACGTCAAGGCTGACAAGCGGGACCCAGAGTTTGTCTTC
TACGACCAGCTGAAGCAAGTGATGAATGCGTACAGAGTCAAGCCGGCCGT
CTTTGACCTGCTCCTGGCTGTTGGCATTGCTGCCTACCTCGGCATGGCCT
ACGTGGCTGTCCAGCACTTCAGCCTCCTCTACAAGACCGTCCAGAGGCTG
CTCGTGAAGGCCAAGACACAGTGAcacagccaccccacagccggagccc
ccgccgctccacagtccctggggccgagcacgagtgagtggacactgccc
cgccgcgggcggccctgcagggacaggggccctctccctccccggcggtg
gttggaacactgaattacagagctttttttctgttgctctccgagactggg
gggggattgtttcttcttttccttgtctttgaacttccttggaggagagc
ttgggagacgtcccggggccaggctacggacttgcggacgagcccccag
tcctgggagccggccgccctcggtctggtgtaagcacacatgcacgatta
aagaggagacgccgggacccccctgcccgatcgcgcgcggcctccgcccac
cgcctcctgccgcaaggggcctggactgcaggcctgacctgctccctgct
ccgtgtctgtcctaggacgtcccctcccgctcccgatggtggcgtggac
atggttatttatctctgctccttcttgcctggaggagggcagtgccagcc

ctggggttctgggattccagccctcctggagcctttgttccccatgtgg
tctcagtgacccgtccccctgacagtgggctcggggagctgcatcaccca
gccttcccccttctccgactgcagggtctgatgtcatcattgacagccttt
gcttcgtgggggcctggcagggcccctgcctccccgacccccgacccact
gcaaatccccgttcccctgcactcctcttctcccagcccatccctccggc
ccctgtgcctctgcggccccagcccagctcccagggccgtcacctgcttg
gccctggcccagctccctgccctgagtcctgagccagtgcctggtgtttc
ctgggctcggtactgggcccccaggccatccaggctttgccacggccagt
tggtcctccctggggaactgggtgcgggtggagtactgggaggcaggagg
tggcccggggaggccttgtggctcctcccctcgctcctcgccctgggcct
cagcttcctcatcaatagaaaggatgtgttcggggtgggggcgtcaggtg
agaacgtttgctgggaaggagaggacttggggcatggcctctggggccac
ccttcctggaactcagagaggaaggtccgggccctcgggaagccttggac
agaaccctccaccccgcagaccaggcgtcgtgtgtgtgtgggagagaagg
aggcccgtgttgagctcagggagaccccggtgtgtccgttctttagcaat
ataacctacccagtgcgtgccgagcaggcttggtggggaagggacttgag
ctgggcaagtcctggcctggcacccgcagccgtctcccttccgtggccca
gggaggtgtttgctgtccgaaggacctgggccggcccatgggagcctggg
gttctgtccagataggaccaggggctctcactttggccaccagttcttcg
gccagcacctctgccctccagaacctgcagcctggaggggtgaggggaca
accacccctctttcctccaggttggcaggggaccctcttctcccgtctgc
cctgcgggttgcccgcctcctccagagacttgcccaagggcccatcacca
ctggcctctgggcacttgtgctgagactctgggacccaggcagctgccac
cttgtcaccatgagagaatttggggagtgcttgcatgctagccagcaggc
tcctgtctgggtgccacggggccagcattttggagggagcttccttcctt
ccttcctggacaggtcgtcatgatggatgcactgactgaccgtctggggc
tcaggctggtgtgggatgcagccggccgatgagaaaataaagccatattg
aatgatcg (SEQ ID NO: 39)

Protein Sequence

>Hs19_11415_28_1_1699.a
MLEEAGEVLENMLKASCLPLGFIVFLPAVLLLVAPPLPAADAAHEFTVYR
MQQYDLQGQPYGTRNAVLNTEARTMAAEVLSRRCVLMRLLDFSYEQYQKA
LRQSAGAVVIILPRAMAAVPQDVVRQFMEIEPEMLAMETAVPVYFAVEDE
ALLSIYKQTQAASASQGSASAAEVLLRTATANGFQMVTSGVQSKAVSDWL
IASVEGRLTGLGGEDLPTIVIVAHYDAFGVAPWLSLGADSNGSGVSVLLE
LARLFSRLYTYKRTHAAYNLLFFASGGKFNYQGTKRWLEDNLDHTDSSL
LQDNVAFVLCLDTVGRGSSLHLHVSKPPREGTLQHAFLRELETVAAHQFP
EVRFSMVHKRINLAEDVLAWEHERFAIRRLPAFTLSHLESHRDGQRSSIM
DVRSRVDSKTLTRNTRIIAEALTRVIYNLTEKGTPPDMPVFTEQMQIQQE
QLDSVMDWLTNQPRAAQLVDKDSTFLSTLEHHLSRYLKDVKQHHVKADKR
DPEFVFYDQLKQVMNAYRVKPAVFDLLLAVGIAAYLGMAYVAVQHFSLLY
KTVQRLLVKAKTQ (SEQ ID NO: 40)

This protein has a transmembrane domain as predicted by SOSUI and TmPred. It also has both a Signal peptide and a transmembrane domain predicted by SMART, suggesting that this is a type I membrane protein with the majority of the protein extracellular.

AW779536

In a comparison of malignant colon samples containing greater than 50% malignant cells in the sample against mixed normal tissues, fragment AW779536 was upregulated 3.7 fold. Enorthern analysis shown in Figure 7 demonstrates that the fragment is expressed in 77% of the tumors and poorly expressed in normal tissue.

Nucleotide sequence of AW779536

TTCTTCCTGTGTTACAATTACCCTGTTTCTGATTACTACAGCCCAACCCGGGCGGA
CACCACCACCATTCTGGCTGCCGGGGCTGGAGTGACCATAGGATTCTGGATCAAC
CATTTCTTCCAGCTTGTATCCAAGCCCGCTGAATCTCTCCCTGTTATTCAGAACAT
CCCACCGNTCACCACCTACATGTTAGNTTTGGGTCTGACCAAATTTGCAGTGGGA
ATTGTGTTGATCCTCTTGGTTCGTCAGCTTGTACAAAATCTCTCACTGCAAGTATT
ATACTCATGGTTCNAGGTNGGTCNCCAGGAACAAGGAGGCCAGGCGGAGACTGG
AGATTGAAGTGCCTTACAAGTTTGTTACCTACACATCTGTTGGCATCTGCGCTAC
AACCTTTGTGCCGATGCTTCACAGGTTTCTGGGATTACCCTGAGTCTCAAACAGT
TGGAAACTAGCCCACTGGACATGAAAGCCAAGACATAGGAAAGTTATTGGTAGG
CAAATCTTGACAACTTATTTTTGTTTAACAACAACAAAAAGTCATACGGCTGTCT
TGCTACT (SEQ ID NO 41)

BLAT searching with this sequence reveals a hypothetical protein predicted by Acembly, Ensembl and Fgenesh++, Hs2_5283_28_1_1143.b with the following nucleotide sequence:

>Hs2_5283_28_1_1143.b
GCTTATGTACAGAAGTACGTCGTGAAGAATTATTTCTACTATTACCTATT
CCAATTTTCAGCTGCTTTGGGCCAAGAAGTGTTCTACATCACGTTTCTTC
Cattcactcactggaatattgacccttatttatccagaagattgatcatcatatgggttttggtgatgtatattg
gccaagtggccaaggatgtcttgaagtggccccgtccctcctcccctccagttgtaaaactggaaaagagactga
tcgctgaatatggaatgccatccacccacgccatggcggccactgccattgccttcaccctccttatctctacta
tggacagataccagtatccatttgtgttgggactggtgatggccgtggtgttttccaccttggtgtgtctcagca
ggctctacactgggatgcatacggtcctggatgtgctgggtggcgtcctgatcaccgcactcctcatcgtcctca
cctaccctgcctggaccttcatcgactgcctggactcggccagccccctcttccccgtgtgtgtcatagttgtgc
cattcttcctgtgttacaattaccctgtttctgattactacagcccaacccgggcggacaccaccaccattctgg
ctgccggggctggagtgaccataggattctggatcaaccatttcttccagcttgtatccaagcccgctgaatctc
tccctgttattcagaacatcccaccactcaccacctacatgttagttttgggtctgaccaaatttgcagtgggaa
ttgtgttgatcctcttggttcgtcagcttgtacaaaatctctcactgcaagtattatactcatggttcaaggtgg
tcaccaggaacaaggaggccaggcggagactggagattgaagtgccttacaagtttgttacctacacatctgttg
gcatctgcgctacaacctttgtgccgatgcttcacaggtttctgggattaccctgagtctcaaacagttggaaac
tagcccactggacatgaaagccaagacataggaaagttattggtaggcaaatcttgacaacttatttttctttaa
caacaacaaaaagtcatacggctgtcttgctactaccagataaatgatgctgctgtgtgaaaggaagaactgtct
catagcggtcattggtcgtccgtggtggttggttgtgctacagttgaacccaggctaaagaccataatccggatc
tttaaaggcacacaccgcgccccccccccccccgcccggcccctgctcctctcgctgttgcacgggctttggatc
tagtcatgggctggcaggaattgtggcctggcttaggaatagctatgagccccactgggttctggagagccagta
gagatggggtgatctgggaggctggaggtagagccttttcttttccgttacaaccttgcctagcatggagttaact
gtgcctggttgggtggtaagatcactctgaaagaaagctcactgtgaagagatgaaaggtggaggcagagctgtg
aggtcatggggaaaagcctgctttccttataagtcctgctgttcatgttggaataaggatctgctcttccttgtt
tccatgcattttgcaggattccaggtaccattaccacactcttctgacccatgaaaccaactggctgctcacaca
tcaccaaacaggttgggggttagccttcagcacaggtggatacatctgggattcactgagattcctgccctctcc
tgcttcctagtggtttgggacaggccctctgcccatcgtcagcagttttttgctttcatacaaacctggaaggca
ctggcatctgcctaggaaagtggatctgtgaagaacagatgaactcaatcctttctggagtctgacaaagaaggg
ataggcttccttgacattgcctgtcctgacaaggcctccctgacattactcctccaatttcacagttaccttctg
taaatctatttttctcatctactgaatagaatcaggcgcccttttttgtcttcccacctcttatctcttggcaattt
taaggggaattaatgcaagaacaactttagtgtctcttgggaaaacaagccaaccaaatacaaaacccattaagc
ctactagggtgagtcctcttaacatgggaaggcgatgattatgcaaacaccggagttccctcctcttcagttcct
aagaataaagaacaggtatcaagaactttctttaaagttagtgtaactatagttaacaaagtatccattgaagtt
tagtgcctgtaggactgagccagtgctttatcaacccaacacatcatcaccatgtgcatactctagaaaaaaaaa
tagcttccttaaaagttacagaggctcttaacgtgttaaaaccgaaaaatcacatttttcttgatttcaaatatg
ttctacggccttactgttgggatgatatttagtatgtaacttagcattccaatttctcaagaatttttaggccgg
gtgcggtggctcatgcctgtaatcccagcactttgggaggccgaggtgggcggaccacgaggtcaggagatcgag
accatcctggctaacacggtaccccgtctctactgaaaatacaaaaaaattagccggacgtggtggagggcgcct
gtagtcccagctactcaggaggctgaggcaggagaatggcgtgaacccggtgagcggagcttgcagtgagccgag
attgcgccactgcactccagcctgggcgacagagcgagactctctcaaaaaaaaaaaaaaagaatttttagcaaa
acatcctgtttttacttaaaattcttctcatatttattatagttagaaggcaaagatcaagatgacctgccgttt
gactgcttttacatcaaactctgcccagtatttgcagcacaactcaggggaagggccttagcttacaggtactcc

cagccttcatctgcccctgcagagcagtggctgtcagccggatgcggcacttttctgtattttcatccacacagc
tgcccagccagagttcgcaacactggatatttacaccaaataattgtggttgacttgtctgaagccagctgacaa
aaggatcagcttttcccacttgtatttttttaaaaagagggattgtgatcattgtcacagagtgggtgctggcctc
tcatatatatgatatatatatatatcattttatatatatatatatatcatatacataatttttactgctgtctctag
ttttaagtcccaacaataggaaggccgatcagctatattgatatatttaaggctgtacttaactaatttgggctg
aggatgaatatatcagccacagcacattaaagaatgagccaaggatttgtcatggttggtcacttttttaaagtat
ttgattactgcaactggagaatgaaaagtgtatattggtgacgccaacctcagtttctgagcactcctgctctgt
ggtgagaatcagacaaaaattcatcggggtgaaaaaggcattacctgattcacacccttgtcttgctagccctct
tccattcatttctcacacagcactttgctctgttaaatcctctctctgtctcagaccattgcttgccccttcaaa
gggtatggttcaggctcctttcaagacatttggagtttctctctgggaaagagagccccctactggtttggctt
cagtctaggtccaccatccctctcgatctggcatcttggagattaatttaaaaggcaagctcaccacaatgtaag
cctatggtctggccaaccttgcttttgggaactgtgacaccaaagcccccaggactatctgcctctccaggagcc
agatagaatgacatgcctttttcctaattgtccacattccacccccaacccactgccactgtgggccaagccatc
catcttgcaatcttcatctaaaacagctctcatttcatgccagttttgctcaaacctgcaccgtcacaagatatt
cagaagatgaaaacgtagaagacacccctgaattaaaaacacttacatagcagtggctggaattactccaaaacg
tgcccagtgatcgcactgtaacatgggatttctcacccaaataggcaactcatgcttcctgagtgtaatcaaag
catgtggtgttttggggccatatgcaccaggtttctattttagaaaccttcagctgtcttgcttatgtactgtat
gtaaatttattctttttaaaaatca cttttatttgattttgacttattaaatgctttaaaagccag
(SEQ ID NO 42)

The amino acid sequence of Hs2_5283_28_1_1143.b is set forth below:

>Hs2_5283_28_1_1143.b
AYVQKYVVKNYFYYYLFQFSAALGQEVFYITFLPFTHWNIDPYLSRRLII
IWVLVMYIGQVAKDVLKWPRPSSPPVVKLEKRLIAEYGMPSTHAMAATAI
AFTLLISTMDRYQYPFVLGLVMAVVFSTLVCLSRLYTGMHTVLDVLGGVL
ITALLIVLTYPAWTFIDCLDSASPLFPVCVIVVPFFLCYNYPVSDYYSPT
RADTTTILAAGAGVTIGFWINHFFQLVSKPAESLPVIQNIPPLTTYMLVL
GLTKFAVGIVLILLVRQLVQNLSLQVLYSWFKVVTRNKEARRRLEIEVPY
KFVTYTSVGICATTFVPMLHRFLGLP (SEQ ID NO 43)

This amino acid sequence is predicted to contain 9 transmembrane domains by SMART and TmPred and 8 transmembrane domains by SOSUI. By contrast, when analyzed by use of the Geneid program, the following gene is identified as being overexpressed in colon tissue.

>chr2_2054
ATGGCGGCCACTGCCATTGCCTTCACCCTCCTTATCTCTACTATGGACAG
ATACCAGTATCCATTTGTGTTGGGACTGGTGATGGCCGTGGTGTTTTCCA
CCTTGGTGTGTCTCAGCAGGCTCTACACTGGGATGCATACGGTCCTGGAT
GTGCTGGGTGGCGTCCTGATCACCGCACTCCTCATCGTCCTCACCTACCC
TGCCTGGACCTTCATCGACTGCCTGGACTCGGCCAGCCCCCTCTTCCCCG
TGTGTGTCATAGTTGTGCCATTCTTCCTGTGTTACAATTACCCTGTTTCT
GATTACTACAGCCCAACCCGGGCGGACACCACCACCATTCTGGCTGCCGG
GGCTGGAGTGACCATAGGATTCTGGATCAACCATTTCTTCCAGCTTGTAT
CCAAGCCCGCTGAATCTCTCCCTGTTATTCAGAACATCCCACCACTCACC
ACCTACATGTTAGTTTTGGGTCTGACCAAATTTGCAGTGGGAATTGTGTT
GATCCTCTTGGTTCGTCAGCTTGTACAAAATCTCTCACTGCAAGTATTAT
ACTCATGGTTCAAGGTGGTCACCAGGAACAAGGAGGCCAGGCGGAGACTG
GAGATTGAAGTGCCTTACAAGTTTGTTACCTACACATCTGTTGGCATCTG
CGCTACAACCTTTGTGCCGATGCTTCACAGGTTTCTGGGATTACCCTGA
SEQ ID NO 44)

This gene encodes a protein having the following predicted structure:
>chr2_2054
MAATAIAFTLLISTMDRYQYPFVLGLVMAVVFSTLVCLSRLYTGMHTVLD
VLGGVLITALLIVLTYPAWTFIDCLDSASPLFPVCVIVVPFFLCYNYPVS
DYYSPTRADTTTILAAGAGVTIGFWINHFFQLVSKPAESLPVIQNIPPLT
TYMLVLGLTKFAVGIVLILLVRQLVQNLSLQVLYSWFKVVTRNKEARRRL
EIEVPYKFVTYTSVGICATTFVPMLHRFLGLP* (SEQ ID NO 45)

When this sequence is analyzed by SOSUI and TmPred it is predicted to possess 7 transmembrane domains. By contrast, analyses by SMART suggests that the protein has 5 transmembrane domains and a signal sequence.

These analyses also indicate that the protein contains a PFAM domain indicating that the protein contains an acid phosphatase domain.

AL531683

In a comparison of malignant colon samples with greater than 50% malignant cells in the sample against mixed normal tissues, fragment AL531683 was found to be upregulated 3.76 fold. The Enorthern analysis shown in Figure 8 demonstrates that the fragment is expressed in 100% of the tumors analyzed and poorly expressed in normal tissue.

The nucleotide sequence of fragment AL53168
CGCCGGCGGTGCGTGTGGGAAGGCGTGGGGTGCGGACCCCGGCCCGACCTCNCC
GTCCCGCCCGCCGCCTTCTGCGTCGCGGGNGCGGGCCGGCGGGGTCCTCTGACGC
GGCAGACAGNCCCTCGCTGTCGCCTCCAGTGGTTGTCGACTTGCGGGCGGCCCCC
CTCCGCGGCGGTGGGGGTGCCGTCCCGCCGGCCCGTCGTGCTGCCCTCTCNNGGG
GGGTTTGCGCGAGCGTCGGCTCCGCCTGGGCCCTTGCGGTGCTCCTGGAGCGCTC
CGGGTTGTCCCTCAGGTGCCCGAGGCCGAACGGTGGTGTGTCGTTCCCGCCCCCG
GCGCCCCCTCCTCCGGTCGCCGCCGCGGTGTCCGCGCGTGGGTCCTGAGGGAGCT
CGTCGGTGTGGGGTTCGAGGCGGTTTGAGTGAGACGAGACGAGAC
(SEQ ID NO 46)

AI202201

In a comparison of malignant colon samples with greater than 50% malignant cells in the sample against mixed normal tissues, fragment AI202201 was upregulated 3.18 fold. Enorthern analysis shown in Figure 9 demonstrates that the fragment is expressed in 77% of the tumors and poorly expressed in normal tissue.

Nucleotide Sequence **AI202201**

ACCCTATAGCTCCTTACGCTGGGAAAGCTGGTTTTTTAAAAAAATAATAATAAAA
TATTTAATCTTATTAAGTGTTCATTTAAAATGCGTAATGCTTTGGAAATAATGGGT
AACAGATAGCGAGAGGATATGTTTATAAAGTGAGCATGTTGGTCCCATTTATAAA
TATATGTATGATTTATAAGCTTTTTTAAAACAAAGCTCAAATTGTTGGTATTTTTC
TAAAATGTGCACAGCTGTATTTTACATGAAGGCTCTTTCTAATGGGTTGTTATACT
GTACTCAACATTTTGGACAGCACATGAAGTCTGCCAATGTACTTAATAAAACATG
ACTTTGTTTATTTAAAGTTTCTTGCTGTGAAAAAGAACTCCCTACCTGTGAGTTCC
TTTATTTATAATTCTTGAAACCAAAATGTATAATGTACAGTTTTCACAACTGTATC
TGCTCTAATA (SEQ ID NO: 47)

AL389942

In a comparison of malignant colon samples with greater than 50% malignant cells in the sample against mixed normal tissues, fragment AL389942 was upregulated 3.83 fold.

Enorthern analysis shown in Figure 10 demonstrates that the fragment is expressed in 55% of the tumors and poorly expressed in normal tissue.

GAAGCTCCAAATGCTCTGGGTTTCAGCTCCTCTGTGCTGTGGACNCTGACTTTGG
CTCAGAACTCCGATTTAGTACAAAAGGCTCATTTTTATTTCAGGGGCACTCTTCCT
AAAGCAAACCTAATAAATGAAATATGGAATTCACAGATACACACACACATTAAA
AAATTAACCTAGTGTATCTGTGAGGAGTAGGCAGAAATTCNCTGTATAAAAGAA
TGCTTCATTTCATAGAGAATTTGTGTTAAGATTCCATTAGATAGTACATTTCTCAA
AGATTTTTGAGGTTGTATTTGCTTTACCAAAACTTGGTTTATGTAAGTGGAAAAA
GCATGTTGCAAAATAACTTGGTGTCTATGATTCAGTTTATGTAAAATAATAAATG
TATGTAGGAATACGTGTGTTGAAAGATGTACATCAATTTGCTAACAATGGTTATC
TCTGACGTGGTGGGATTTGAGATGTGTTTTTCTTTTTGGTTGTATTTTTCTCTATTG
TTTGACTTA (SEQ ID NO: 48)

EXAMPLE 5

Through the same collaboration described in example 1 and using the Celera Database the following DNA sequences were identified that are overexpressed in malignant colon tissues as well as some other cancers.

bs243ms232-222
The bs243ms232-222 gene, having the sequence set forth below, was found to be overexpressed in colon cancer.

bs243ms232-222
GATCCTGGGACCCCTGGGCCGTGCCTGCCCTCCACCTTGAGTGCCATACTCCCAACAGCTCC
AGGTACCCACCGGGGGATGTGCCTGCTCAGGAAACCTCTTTGCTCCACACAGCATGGGGCTT
CAGCTGCTGGCCCAAGGCCAGGAGCGCTGGGTTCTGCAGCAGGGCTCAGCCTCAGGGGCGTT
A (SEQ ID NO.: 49)

We analyzed this sequence using the Celera database and found that it corresponds to the 3'UTR of the hypothetical protein Hs16_15516_28_2_1402.a predicted by the Acembly program, C16000171 predicted by the FGENESH program, chr16_148 predicted by the GeneID program and NT_015360.30 predicted by the GeneScan program. The Hs16_15516_28_2_1402a sequence is set forth below. This sequence contains a 5' and 3' UTRs, which are contained therein:

>Hs16_15516_28_2_1402.a
ccctcccgcgtccggccgcgcccgtcctcctggctgcagagagactaccg
gccaccgccgccgccgccgcgagctgtccctgcggcgcgtctgcctt
ggcggagccgaccgcagtgcgctcaggcgtccggtgcgtccccagcctcc
gccccggcgcgggggcgacggactcgcgcgtgcgcagcgccggaggggcg
cgggctgggacccctagccagcgcgtgcgccgatcgagcgcagggcgat
gggtgggcgccgggcgccgggcgccaggcagtgatgggccttcccgcgct
gcggccccactgaggaggaggctcggggacagcaggagcacgggctgccc
gcgcggtgcggaccATGGCGTTCCTGGCCGGGCCGCGCCTGCTGGACTGG
GCCAGCTCGCCGCCGCACCTGCAGTTCAATAAGTTCGTGCTGACCGGGTA
CCGGCCCGCCAGCAGCGGCTCGGGCTGCCTGCGCAGCCTCTTCTACCTGC
ACAACGAACTGGGCAACATCTACACGCACGGGCTGGCCCTGCTGGGCTTC

CTGGTGCTGGTGCCAATGACCATGCCCTGGGGTCAGCTGGGCAAGGATGG
CTGGCTGGGAGGCACACATTGCGTGGCCTGCCTTGCACCCCCTGCAGGCT
CCGTGCTCTATCACCTCTTTATGTGCCACCAAGGGGGCAGCGCTGTGTAC
GCCCGGCTCCTCGCCCTGGACATGTGTGGGGTCTGCCTTGTCAACACCCT
TGGGGCCCTGCCCATCATCCACTGCACCCTGGCCTGCAGGCCCTGGCTGC
GCCCGGCTGCCCTGGTGGGCTACACTGTGTTGTCGGGTGTGGCCGGCTGG
CGTGCTCTCACCGCCCCCTCCACCAGTGCTCGGCTCCGGGCATTTGGATG
GCAGGCTGCTGCCCGCCTACTGGTATTTGGGGCCCGGGGAGTGGGTCTGG
GTTCAGGGGCTCCAGGCTCCCTGCCCTGCTACCTGCGCATGGACGCACTG
GCGCTGCTTGGGGGACTGGTAAATGTAGCCCGTCTGCCCGAGCGCTGGGG
ACCTGGCCGCTTTGACTACTGGGGCAACTCCCACCAGATCATGCACCTGC
TGAGCGTGGGCTCCATCCTGCAGCTGCACGCCGGCGTCGTGCCCGACCTG
CTCTGGGCTGCCCACCACGCCTGTCCCCGGGACTGAgctgccatgccagc
ctgcccacagcagcctcctagagttagcaacaccaggtgttcctcccaac
tcgtctgcaaggggctggctccttggatgcttccagctcatgagatgtct
cagcaggagccctgttcacccgttcttccctgtggactgacctcttccac
ccacgccgtggcgctccaacttccttccctgccttttccctccaagctcc
tattttactgtgtcagctggaaggaaaccttccctcttgggacctcttt
accctctgtgacctgtggggttagaccagagagggactctggggtcacgt
cttgctctgagagttcaagtcctgccaggccgccagcccagagcctcctc
accctatcctgttcctcccaccaggcctgtggccagtcttcctgatctcc
atctttctgccctgcataccagccctcccagcagccacaagcttgcccgc
cctggctccctctgcccagagactatggagtaaggcattcaggacaaaag
gaccaaggggcgtggacccgtcttgtaccagctggccacaggcacaagg
gctgcagctgcttcttccaggaaactgacacagggagctcagcggcctca
gatcctgggacccctgggccgtgcctgccctccaccttgagtgccatact
cccaacagctccaggtacccaccgggggatgtgcctgctcaggaaacctc
tttgctccacacagcatggggcttcagctgctggcccaaggccaggagcg
ctgggttctgcagcagggctcagcctcaggggcgttaagaccctggatga
catcaataaagggacaggaagggccatgttgccacatgagcaagcttggg
tgctcccaaggttcaaatacttttattagacacggccaggcagagaaga
ccatgggagttcccgaggggccccagctttcaagggcgacgggagagaca
caggataaaaggttaaaagtgcagaggcagagtctggggctcaggttggg
tctagggtgtcctcaaacaggctgaggaggttccgaggctcaaaggaggg
gaaggagccccgaggaggctctgagttgatgtcacttaggtccagggcat
ccctgggaggagagagtagtgacactcaggatccaaaagctagccctgcc
caccccagcccctggacctgcttacctgggtgtgcacctgctccgggggg
tggaggtgctccccacagtccgggccaggacagcctcaggggagagtgaa
ggcctgcaggagggcaggcgagacaaggagggtgtccagggctagggagt
gccggatgaaaccagctctgtccctgtgcaggctccaggctcccgcctga
caaacaggcagggagccacagtcagggacaataaaaacttggtgcactct
gaaagcagcacttggacagccttcaaagtccttccatctggctgcactcc
aaggccccctctgtccttttcagaacacatggacttggaggcagatttga
aataaacttttagtaaatgtaa (SEQ ID NO. 50)

HS16_15516_28_2_1402.a codes for the following protein:

>Hs16_15516_28_2_1402.a
MAFLAGPRLLDWASSPPHLQFNKFVLTGYRPASSGSGCLRSLFYLHNELG
NIYTHGLALLGFLVLVPMTMPWGQLGKDGWLGGTHCVACLAPPAGSVLYH
LFMCHQGGSAVYARLLALDMCGVCLVNTLGALPIIHCTLACRPWLRPAAL
VGYTVLSGVAGWRALTAPSTSARLRAFGWQAAARLLVFGARGVGLGSGAP
GSLPCYLRMDALALLGGLVNVARLPERWGPGRFDYWGNSHQIMHLLSVGS

ILQLHAGVVPDLLWAAHHACPRD (SEQ ID NO. 51)

This protein was analyzed using various computerized analyses methods which revealed that the protein apparently contains 3 transmembrane domain (as determined by SMART, 2 by SOSUI), and 5 by TmPred. This analysis indicates SMART that this protein contains an area of low complexity (the pink region). The predicted protein structure is shown schematically below.

Further analysis of this protein sequence revealed that the 3' UTR of this sequence overlaps with the 3' UTR of a different protein (membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase (PKMYT1)). The EST in GeneLogic contains sequences that could be from either gene. The expression data for this protein suggests that this gene is also upregulated in 30% of breast, colon, prostate, rectum and stomach malignancies. Based thereon, this gene and corresponding protein may prove to be a suitable target for breast, colon, rectal and stomach malignancies.

The Enorthern of the EST representing bs243ms232-222 is contained in Figure 11.

Additionally, when the bs243ms232-222 sequence was searched against the PFAM motif database, (both through the SMART database and the Profile Scan Servers), amino acids 33-259 show homology to UPF0073 (Uncharacterized protein family (Hly-III / UPF0073)) with an E value of 4.8 e-08 (SMART) and 2.8 e-08 (Profile). We have named this the full gene **CHEM1** (Colon Hemolysin containing, Expressed in other Malignancies), based on its expression in malignancies other than colon cancer.

To confirm the data from the GeneExpress program, the inventors performed PCR validation. In short, intron-spanning primers were designed in order to investigate expression in cDNAs from multiple tissue panels obtained from Clontech using GAPDH as an internal control. As shown in Figures 12-16, the CHEM1 message is overexpressed in malignant colon and prostate as compared to normal organs.

Further analysis of the bs243ms232-222 sequence also suggests that there may be an alternatively spliced transcript. This predicted splice variant, UPF0073.5.b is set forth below. UPF0073.5c, d, and e are alternatively spliced transcripts without changes to the coding sequence and are not depicted.

>UPF0073.5.b
ctggcgtcccctcccgcgtccggccgcgcccgtcctcctggctgcagaga
gactaccggccaccgccgccgccgccgccgcgagctgtccctgcggcgcg
tctgccttggcggagccgaccgcagtgcgctcaggcgtccggtgcgtccc
cagcctccgccccggcgcgggggcgacggactcgcgcgtgcgcagcgccg
gaggggcgcgggctgggacccccctagccagcgcgtgcgccgatcgagcgc
agggcgatgggtgggcgccgggcgccgggcgccaggcagtgatgggcctt
cccgcgctgcggccccactgaggaggaggctcggggacagcaggagcacg
ggctgcccgcgcggtgcggaccATGGCGTTCCTGGCCGGGCCGCGCCTGC

TGGACTGGGCCAGCTCGCCGCCGCACCTGCAGTTCAATAAGTTCGTGCTG
ACCGGGTACCGGCCCGCCAGCAGCGGCTCGGGCTGCCTGCGCAGCCTCTT
CTACCTGCACAACGAACTGGGCAACATCTACACGCACGGCTCCGTGCTCT
ATCACCTCTTTATGTGCCACCAAGGGGGCAGCGCTGTGTACGCCCGGCTC
CTCGCCCTGGACATGTGTGGGGTCTGCCTTGTCAACACCCTTGGGGCCCT
GCCCATCATCCACTGCACCCTGGCCTGCAGGCCCTGGCTGCGCCCGGCTG
CCCTGGTGGGCTACACTGTGTTGTCGGGTGTGGCCGGCTGGCGTGCTCTC
ACCGCCCCCTCCACCAGTGCTCGGCTCCGGGCATTTGGATGGCAGGCTGC
TGCCCGCCTACTGGTATTTGGGGCCCGGGGAGTGGGTCTGGGTTCAGGGG
CTCCAGGCTCCCTGCCCTGCTACCTGCGCATGGACGCACTGGCGCTGCTT
GGGGGACTGGTAAATGTAGCCCGTCTGCCCGAGCGCTGGGGACCTGGCCG
CTTTGACTACTGGGGCAACTCCCACCAGATCATGCACCTGCTGAGCGTGG
GCTCCATCCTGCAGCTGCACGCCGGCGTCGTGCCCGACCTGCTCTGGGCT
GCCCACCACGCCTGTCCCCGGGACTGAgctgccatgccagcctgcccaca
gcagcctcctagagttagcaacaccaggtgttcctcccaactcgtctgca
aggggctggctccttggatgcttccagctcatgagatgtctcagcaggag
ccctgttcacccgttcttccctgtggactgacctcttccacccacgccgt
ggcgctccaacttccttcccctgccttttccctccaagctcctatttact
gtgtcagctggaaggaaaccttcccctcttgggacctctttaccctctgt
gacctgtggggttagaccagagagggactctggggtcacgtcttgctctg
agagttcaagtcctgccaggccgccagcccagagcctcctcaccctatcc
tgttcctcccaccaggcctgtggccagtcttcctgatctccatctttctg
ccctgcataccagccctcccagcagccacaagcttgcccgccctggctcc
ctctgcccagagactatggagtaaggcattcaggacaaaaggaccaaggg
ggcgtggacccgtcttgtaccagctggccacaggcacaagggctgcagct
gcttcttccaggaaactgacacagggagctcagcggcctcagatcctggg
acccctgggccgtgcctgccctccaccttgagtgccatactcccaacagc
tccaggtacccaccgggggatgtgcctgctcaggaaacctctttgctcca
cacagcatggggcttcagctgctggcccaaggccaggagcgctgggttct
gcagcagggctcagcctcaggggcgttaagaccctggatgacatcaataa
agggacaggaagggccatgttgccacatgagcaagcttgggtgctcccaa
ggttcaaatacttttattagacacggccaggcagagaagaccatgggag
ttcccgaggggccccagctttcaagggcgacgggagagacacaggataaa
aggttaaaagtgcagaggcagagtctggggctcaggttgggtctagggtg
tcctcaaacaggctgaggaggttccgaggctcaaaggaggggaaggagcc
ccgaggaggctctgagttgatgtcacttaggtccagggcatccctgggag
gagagagtagtgacactcaggatccaaaagctagccctgcccaccccagc
ccctggacctgcttacctgggtgtgcacctgctccgggggggtggaggtgc
tccccacagtccgggccaggacagcctcaggggagagtgaaggcctgcag
gagggcaggcgagacaaggagggtgtccagggctagggagtgccggatga
aaccagctctgtccctgtgcaggctccaggctcccgcctgacaaacaggc
agggagccacagtcagggacaataaaaacttggtgcactctgaaagcagc
acttggacagccttcaaagtccttccatctggctgcactccaaggccccc
tctgtccttttcagaacacatggacttggaggcagatttgaaataaactt
ttagtaaatgtaagcctt (SEQ ID NO.: 52)

The amino acid sequence for this splice variant is shown below:

>UPF0073.5.b
MAFLAGPRLLDWASSPPHLQFNKFVLTGYRPASSGSGCLRSLFYLHNELG
NIYTHGSVLYHLFMCHQGGSAVYARLLALDMCGVCLVNTLGALPIIHCTL
ACRPWLRPAALVGYTVLSGVAGWRALTAPSTSARLRAFGWQAAARLLVFG
ARGVGLGSGAPGSLPCYLRMDALALLGGLVNVARLPERWGPGRFDYWGNS

HQIMHLLSVGSILQLHAGVVPDLLWAAHHACPRD (SEQ ID NO.53)

Analysis of this protein sequence using protein analysis programs suggests that this protein may have one or three transmembrane domains. Although SMART does not predict the hemolysin domain in the shorter version, Profile predicts the UPF0073 domain with an E value of 4.9e-06.

Murine Homolog of CHEM1 Gene

Murine CHEM1

IDEC proposes to develop an animal model of CHEM1 to test potential therapeutics that target the CHEM1 gene or protein. Accordingly, we investigated whether a murine homologue of CHEM1 exists. Using the protein BLAST database, the following murine homologue was identified:

>gi|12963841|ref|NP_076313.1| RIKEN cDNA 1500004C10 [Mus musculus]
MAFLTGPRLLDWASSPPHLQFNKFVLTGYRPASSGSGCLRSLFYLHNELGNIYTHGLALLGF
LVLVPMTMPWSQLGKDGWLGGTHCVACLVPPAASVLYHLFMCHQGGSPVYTRLLALDMCGVC
LVNTLGALPIIHCTLACRPWLRPAALMGYTALSGVAGWRALTAPSTSARLRAFGWQAGARLL
VFGARGVGLGSGAPGSLPCYLRMDALALLGGLVNVARLPERWGPGRFDYWGNSHQIMHLLSV
GSILQLHAGVVPDLLWAAHHACPPD(SEQ ID NO.54)

The nucleotide sequence of the murine CHEM1 protein is:

>gi|12963840|ref|NM_023824.1| Mus musculus RIKEN cDNA 1500004C10 gene (1500004C10Rik), mRNA
ATGCACTGAGCTCCGACCTGGGGTTGCCAGCTTTCTCTCCCTTGCGGGGGCGTCGAACTCGCGCGTGCGC
AGCGCGTGAGGGAAGGGGGCCGGGACCTCCTTGCTGACCCGGGCAGGGCCACCGGATAGCCGGAGGTGAA
TCGGGATGAGCTTCCCAGCGCTGCAGCTCCACTGAGAAGGAAGCCCAGGCGCAGAGGGTCGCCGGTCGGC
CGCAGTGCGTGAGGCCATGGCATTCCTGACCGGGCCTCGTCTCCTGGACTGGGCTAGCTCGCCGCCGCAC
CTGCAGTTCAATAAGTTCGTATTAACCGGCTACCGGCCGGCGAGCAGCGGCTCGGGCTGTCTGCGCAGCC
TTTTCTACCTACACAACGAGCTGGGCAACATCTACACACACGGGCTAGCCCTGCTGGGCTTCCTGGTGTT
GGTGCCAATGACCATGCCCTGGAGTCAGCTGGGCAAGGATGGCTGGCTAGGAGGTACACACTGTGTGGCT
TGCCTGGTGCCCCCTGCAGCCTCTGTGCTGTATCACCTCTTCATGTGCCACCAAGGAGGCAGTCCTGTGT
ACACCCGGCTCCTTGCCTTGGATATGTGTGGAGTCTGCCTTGTCAACACCCTTGGAGCCCTGCCCATCAT
CCATTGCACTCTGGCCTGCAGACCGTGGCTTCGCCCAGCTGCCCTGATGGGTTACACTGCACTGTCAGGT
GTAGCCGGCTGGAGAGCTCTCACTGCCCCCTCCACCAGTGCCCGGCTTCGAGCCTTTGGTTGGCAAGCTG
GGGCCCGCCTGCTGGTGTTTGGGGCCCGTGGAGTGGGGCTGGGCTCAGGGGCTCCAGGCTCTCTGCCCTG
CTACCTGCGCATGGACGCACTGGCTCTGCTTGGAGGGCTGGTGAATGTGGCACGCCTGCCAGAGCGGTGG
GGGCCTGGTCGCTTCGACTACTGGGGCAACTCCCACCAGATCATGCACTTGCTGAGTGTGGGCTCCATCC
TCCAGCTCCATGCTGGGGTTGTGCCTGACCTGCTCTGGGCTGCACACCATGCCTGTCCCCCAGACTGAGC
TGCCTCCTAGCTGCCAAACTGGCTTGCCCACAGCTTCCTGGACAAATTCCACCACCTTTCCTCCTACTGG
TCTGCAAGGGGCTGGTTCCCTGGAAGAACCAGCACATGGGACTTCCTAGCTGGGAGACCATTCTTCATTC
TTCCCCATGGATTCACTTCTTGCATCCAGGCCTTCAAACCCCAGCTTCCACTTTCCTTGCCATCTTCCCT
CCTGGGCATTGTTTTGCTGTCATTAGAAGGAAACCATTTTTTTTTTTCCCAATTTACCCTGTTTAACCTG
TGAGAGTCTCTGACAGTTGAGTCCTGCCAACTTACCAAGCCTCCAGCCCAGAACCACTACCCCTATGTTG
CTGCTCCCATACATAACTACACCTCCTGCTCCTGGATTCTTGAGCTAGCCACTCTGACCCTGCTTCCTGA
CCTCCATCTCCCTGCTCTGCATGTCAAACCTCTCAGCAGCCAGAATTTTGCTGTTCCTGTCATTCCTGCA
GTGAGGATGCAGAGGAGTGGGACCAGGCTTCTCTCAGAGCCAAGTGGACATTGGTCCTGCTTGTATCATC
TGGCCAGGAGACAGGAGGGGAACTGCTGCTTTTCCTAGGCAACAGGCACAGCTGTGGAATGGAGGTGTTG
GATTCGGGCTTCACTGGACCAAGGACTCAGCTCTTCAGTGCCATGGTCTGACTGACCTGCCTACCAGAGA
CTTGTCTGCTCAGGAAATCTCTATACAGTGGGTGGCTCCAGCCTGCTGGCCCAAGGGTACTGACTCGCAG
CCAGATCATCCCAAAGGCCCAAGACCCTAGGCAACATCAATAAAGGGACAAGAAGAGCTATGCTGCCACA
TGAGCAACCTTGGGTGTTCCCAAGACGCATTACTTTTTATTAGACACGGAAGTTTCAGGGGAGAGGTGGG
CAAGACGGTCAGAGGTTTAAAAGCACCAAGGCTGGCTGGGCCTGTGCTCAGGCTGGGTCTAGGGAGTCCT
CAAACAGGCTGAGGAGGTTCCTTGGCTCAAAGGTGGGGCAGGGACCTCTTGGAGGCTCTGAGTCCACATC

AGTTAGGTCCAGGGCATCCCTTGGGGGAGGAAGAAGAAGAAAAAAAAAAAAAAAAAAAAGGCCACA
(SEQ ID NO.: 55)

Based on this information, animal models will be developed using antibodies that target mouse CHEM1. These antibodies will be naked or may be conjugated to an effector moiety, e.g., a radionucleotide to test the ability of CHEM1 as an appropriate target for human cancer therapy, especially for treatment of human colon cancer and potentially also breast, rectal, stomach and prostate cancer, as this protein seems to be overexpressed in these tissues.

EXAMPLE 6

IDENTIFICATION OF GENE UPREGULATED IN COLON CANCER

Using a DNA referred to as Ly6G6C revealed that a DNA fragment NM_021246 appears to be 5-fold up-regulated as shown by hybridization in the malignant colon compared with mixed normal samples, greater than 3-fold up-regulated compared with normal kidney, liver and lung, and greater than 2-fold up-regulated in all other tissues.

NM_021246
AACCGAATGCGGTGCTACAACTGTGGTGGAAGCCCCAGCAGTTCTTGCAAAGAGGCCGTGAC
CACCTGTGGCGAGGGCAGACCCCAGCCAGGCCTGGAACAGATCAAGCTACCTGGAAACCCCC
CAGTGACCTTGATTCACCAACATCCAGCCTGCGTCGCAGCCCATCATTGCAATCAAGTGGAG
ACAGAGTCGGTGGGAGACGTGACTTATCCAGCCCACAGGGACTGCTACCTGGGAGACCTGTG
CAACAGCGCCGTGGCAAGCCATGTGGCCCCTGCAGGCATTTTGGCTGCAGCAGCTACCGCCC
TGACCTGTCTCTTGCCAGGACTGTGGAGCGGATAGGGGGAGTAGGAGTAGAGAAGGGAACAA
GGGAGCAAGGGAACAAGGGACATCTGAACATCT
(SEQ ID NO.: 56)

This is substantiated by the Enorthern results contained in Figure 17.

The Enothern results in Figure 17 indicate that this fragment is up-regulated in colon and rectal malignancies. Accordingly, this gene may be targeted for the treatment of colon or rectal cancer. A search of commercial databases reveals that NM_021246 is apparently part the Ly6G6D gene is set forth below:

>Ly6G6D message
cccatggcagtcttattcctcctcctgttcctatgtggaactccccaggc
tgcagacaacatgcaggccatctatgtggccttgggggaggcagtagagc
tgccatgtccctcaccacctactctacatggggacgaacacctgtcatgg
ttctgcagccctgcagcaggctccttcaccaccctggtagcccaagtcca
agtgggcaggccagccccagaccctggaaaaccaggaagggaatccaggc
tcagactgctggggaactattctttgtggttggagggatccaaagaggaa
gatgccgggcggtactggtgcgctgtgctaggtcagcaccacaactacca
gaactggagggtgtacgacgtcttggtgctcaaaggatcccagttatctg
caagggctgcagatggatccccctgcaatgtcctcctgtgctctgtggtc
cccagcagacgcatggactctgtgacctggcaggaagggaagggtcccgt
gaggggccgtgttcagtccttctggggcagtgaggctgccctgctcttgg
tgtgtcctggggaggggctttctgagcccaggagccgaagaccaagaatc
atccgctgcctcatgactcacaacaaaggggtcagctttagcctggcagc

ctccatcgatgcttctcctgccctctgtgccccttccacgggctgggaca
tgccttggattctgatgctgctgctcacaatgggccagggagttgtcatc
ctggccctcagcatcgtgctctggaggcagagggtccgtggggctccagg
cagaggaaaccgaatgcggtgctacaactgtggtggaagccccagcagtt
cttgcaaagaggccgtgaccacctgtggcgagggcagaccccagccaggc
ctggaacagatcaagctacctggaaacccccccagtgaccttgattcacca
acatccagcctgcgtcgcagcccatcattgcaatcaagtggagacagagt
cggtgggagacgtgacttatccagcccacagggactgctacctgggagac
ctgtgcaacagcgccgtggcaagccatgtggcccctgcaggcatttggc
tgcagcagctaccgccctgacctgtctcttgccaggactgtggagcggat
agggggagtaggagtagagaagggaacaagggagcaagggaacaagggac
atctgaacatctaatgtgagaagagaaacatccttctgtgagtcattaaa
atctatgaaccactct (SEQ ID NO.: 57)

The amino acid sequence for Ly6G6D is set forth below:

>Ly6G6D protein
MAVLFLLLFLCGTPQAADNMQAIYVALGEAVELPCPSPPTLHGDEHLSWF
CSPAAGSFTTLVAQVQVGRPAPDPGKPGRESRLRLLGNYSLWLEGSKEED
AGRYWCAVLGQHHNYQNWRVYDVLVLKGSQLSARAADGSPCNVLLCSVVP
SRRMDSVTWQEGKGPVRGRVQSFWGSEAALLLVCPGEGLSEPRSRRPRII
RCLMTHNKGVSFSLAASIDASPALCAPSTGWDMPWILMLLLTMGQGVVIL
ALSIVLWRQRVRGAPGRGNRMRCYNCGGSPSSSCKEAVTTCGEGRPQPGL
EQIKLPGNPPVTLIHQHPACVAAHHCNQVETESVGDVTYPAHRDCYLGDL
CNSAVASHVAPAGILAAAATALTCLLPGLWSG (SEQ ID NO.: 58)

Analysis of the Ly6G6D protein sequence using the SMART program suggests that this protein has two transmembrane domains and an Ig domain, suggesting that this protein is a cell surface protein.

EXAMPLE 7
Identification of Colon-Cancer Associated Gene AI821606

FLJ32334

Fragment AI821606 set forth below, also was shown to be upregulated in colon, pancreas and rectal malignancies. This is supported by the Enorthern results in Figure 18.

AI821606
TTCCTCGGAGGGGCCGTGGTGAGTCTCCAGTATGTTCGGCCCAGCGCTCTTCGCACCCTTCT
GGACCAAAGCGCCAAGGACTGCAGCCAGGAGAGAGGGGGCTCACCTCTTATCCTCGGCGACC
CACTGCACAAGCAGGCCGCTCTCCCAGACTTAAAATGTATCACCACTAACCTGTGAGGGGGA
CCCAATCTGGACTCCTTCCCCGCCTTGGGACATCGCAGGCCGGGAAGCAGTGCCCGCCAGGC
CTGGGCCAGGAGAGCTCCAGGAAGGGCACTGAGCGCTGCTGGCGCGAGGCCTCGGACATCCG
CAGGCACCAGGGAAAGTCTCCTGGGGCGATCTGTAAAT (SEQ ID NO.: 59)

A database search revealed that AI821606 is in the 3'UTR of predicted genes corresponding to both strands of a chromosome. Based thereon, the following structure for this fragment is predicted.

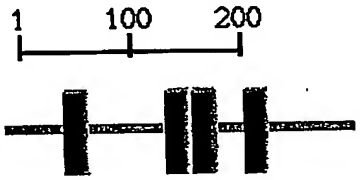
FLJ32334
>ENST00000267803
gcttccagcggacggcagcgcgcgagcattgccccccctgcaccacctca
ccaagATGGCTACTTTGGGACACACATTCCCCTTCTATGCTGGCCCCAAG
CCAACCTTCCCGATGGACACCACTTTGGCCAGCATCATCATGATCTTTCT
GACTGCACTGGCCACGTTCATCGTCATCCTGCCTGGCATTCGGGGAAAGA
CGAGGCTGTTCTGGCTGCTTCGGGTGGTGACCAGCTTATTCATCGGGGCT
GCAATCCTGGGGACCCCCGTGCAGCAGCTGAATGAGACCATCAATTACAA
CGAGGAGTTCACCTGGCGCCTGGGTGAGAACTATGCTGAGGAGTATGCAA
AGGCTCTGGAGAAGGGGCTGCCAGACCCTGTGTTGTACCTAGCTGAGAAG
TTCACTCCAAGAAGCCCATGTGGCCTATACCGCCAGTACCGCCTGGCGGG
ACACTACACCTCAGCCATGCTATGGGTGGCATTCCTCTGCTGGCTGCTGG
CCAATGTGATGCTCTCCATGCCTGTGCTGGTATATGGTGGCTACATGCTA
TTGGCCACGGGCATCTTCCAGCTGTTGGCTCTGCTCTTCTTCTCCATGGC
CACATCACTCACCTCACCCTGTCCCCTGCACCTGGGCGCTTCTGTGCTGC
ATACTCACCATGGGCCTGCCTTCTGGATCACATTGACCACAGGACTGCTG
TGTGTGCTGCTGGGCCTGGCTATGGCGGTGGCCCACAGGATGCAGCCTCA
CAGGCTGAAGGCTTTCTTCAACCAGAGTGTGGATGAAGACCCCATGCTGG
AGTGGAGTCCTGAGGAAGGTGGACTCCTGAGCCCCCGCTACCGGTCCATG
GCTGACAGTCCCAAGTCCCAGGACATTCCCCTGTCAGAGGCTTCCTCCAC
CAAGGCATACTGTAAGGAGGCACACCCCAAAGATCCTGATTGTGCTTTAt
aacattcctccccgtggaggccacctggacttccagtctggctccaaacc
tcattggcgccccataaaaccagcagaactgccctcagggtggctgttac
cagacacccagcaccaatctacagacggagtagaaaaaggaggctctata
tactgatgttaaaaaacaaaacaaaacaaaaagccctaagggactgaaga
gatgctgggcctgtccataaagcctgttgccatgataaggccaagcaggg
gctagcttatctgcacagcaacccagcctttccgtgctgccttgcctctt
caagatgctattcactgaaacctaacttcacccccataacaccagcaggg
tgggggttacatatgattctcctatggtttcctctcatccctcggcacct
cttgttttccttttttcctgggttccttttgttcttcctttacttctccag
cttgtgtggccttttggtacaatgaaagacagcactggaaaggaggggaa
accaaacttctcatcctaggtctaacattaaccaactatgccacattctc
tttgagcttcagttcccaaatttgctacataagattgcaagacttgccaa
gaatcttgggatttatctttctatgccttgctgacacctaccttggccct
caaacaccacctcacaagaagccaggtgggaagttagggaatcaactcca
aaacgctattccttcccaccccactcagctgggctagctgagtggcatcc
aggacgggggagtgggtgacctgcctcatcactgccacctaacgtccccc
tggggtggttcagaaagatgctagctctggtagggtccctccggcctcac
tagagggcgcccctattactctggagtcgacgcagagaatcaggtttcac
agcactgcggagagtgtactaggctgtctccagcccagcgaagctcatga
ggacgtgcgaccccggcgcggagaagccatgaaaattaatgggaaaaaca
gtttttaaaaaacaaaagaaaaaaaggtttatttacagatcgccccagga
gactttccctggtgcctgcggatgtccgaggcctcgcgccagcagcgctc
agtgcccttcctggagctctcctggcccaggcctggcgggcactgcttcc
cggcctgcgatgtcccaaggcggggaaggagtccagattgggtccccctc
acaggttagtggtgatacattttaagtctgggagagcggcctgcttgtgc
agtgggtcgccgaggataagaggtgagccccctctctcctggctgcagtc
cttggcgctttggtccagaagggtgcgaagagcgctgggccgaacatact
ggagactcaccacggcccctccgaggaagaggcacaggacgcctgtggcg
gtggggatcgaaagaaaggagggcatgtggagtcagggctatgttgccca
ggctggtctcgaactctggcctcaaacgaccttcctgcctcgacctccca
aagtgctgggattacaggcgtgatgcccgggccttcttccatcttttgga
gcctacccccttgtgttacctcccgccacacacctctaatctgaattacat

gaaacacggcaagacaccaaaccctctgagccccccactttcatctgt
aaaatggtcataacagtgcctgtttctgcgaactattgagaggggcaaat
agggtaatagatgtgaattcattctgtaaactgg (SEQ ID NO.: 60)

The predicted coding sequence for ENST00000267803 is set forth below:

>ENST00000267803
MATLGHTFPFYAGPKPTFPMDTTLASIIMIFLTALATFIVILPGIRGKTR
LFWLLRVVTSLFIGAAILGTPVQQLNETINYNEEFTWRLGENYAEEYAKA
LEKGLPDPVLYLAEKFTPRSPCGLYRQYRLAGHYTSAMLWVAFLCWLLAN
VMLSMPVLVYGGYMLLATGIFQLLALLFFSMATSLTSPCPLHLGASVLHT
HHGPAFWITLTTGLLCVLLGLAMAVAHRMQPHRLKAFFNQSVDEDPMLEW
SPEEGGLLSPRYRSMADSPKSQDIPLSEASSTKAYCKEAHPKDPDCAL
(SEQ ID NO. 61)

We analyzed the protein using SMART and predict that the protein contains several transmembrane domains and a signal sequence, as depicted schematically below:



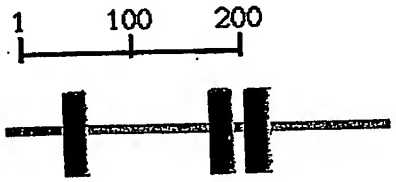
Based on a sequence contain on the opposite strand of the chromosome, the following gene sequence is predicted.

>chr15.41.013.a
ATGACCCTGTGGAACGGCGTACTGCCTTTTTACCCCCAGCCCCGGCATGC
CGCAGGCTTCAGCGTTCCACTGCTCATCGTTATTCTAGTGTTTTTGGCTC
TAGCAGCAAGCTTCCTGCTCATCTTGCCGGGGATCCGTGGCCACTCGCGC
TGGTTTTGGTTGGTGAGAGTTCTTCTCAGTCTGTTCATAGGCGCAGAAAT
TGTGGCTGTGCACTTCAGTGCAGAATGGTTCGTGGGTACAGTGAACACCA
ACACATCCTACAAAGCCTTCAGCGCAGCGCGCGTTACAGCCCGTGTCCGT
CTGCTCGTGGGCCTGGAGGGCATTAATATTACACTCACAGGGACCCCAGT
GCATCAGCTGAACGAGACCATTGACTACAACGAGCAGTTCACCTGGCGTC
TGAAAGAGAATTACGCCGCGGAGTACGCGAACGCACTGGAGAAGGGGCTG
CCGGACCCAGTGCTCTACCTGGCGGAGAAGTTCACACCGAGTAGCCCTTG
CGGCCTGTACCACCAGTACCACCTGGCGGGACACTACGCCTCGGCCACGC
TATGGGTGGCGTTCTGCTTCTGGCTCCTCTCCAACGTGCTGCTCTCCACG
CCGGCCCCGCTCTACGGAGGCCTGGCACTGCTGACCACCGGAGCCTTCGC
GCTCTTCGGGGTCTTCGCCTTGGCCTCCATCTCTAGCGTGCCGCTCTGCC
CGCTCCGCCTAGGCTCCTCCGCGCTCACCACTCAGTACGGCGCCGCCTTC
TGGGTCACGCTGGCAACCGGTGAGGACCGAGAGAATGGGCCCCGGGGGCT
AAGGGTGGAGACAGGATTCACACCGGGCGTCCTGTGCCTCTTCCTCGGAG
GGGCCGTGGCCGGGAAGCAGTGCCCGCCAGGCCTGGGCCAGGAGAGCTCC
AGGAAGGGCACTGAGCGCTGCTGGCGCGAGGCCTCGGACATCCGCAGGCA
CCAGGGAAAGTCTCCTGGGGCGATCTGTAAA (SEQ ID NO. 62)

This sequence is predicted to encode the following protein:

>chr15.41.013.a
MTLWNGVLPFYPQPRHAAGFSVPLLIVILVFLALAASFLLILPGIRGHSR
WFWLVRVLLSLFIGAEIVAVHFSAEWFVGTVNTNTSYKAFSAARVTARVR
LLVGLEGINITLTGTPVHQLNETIDYNEQFTWRLKENYAAEYANALEKGL
PDPVLYLAEKFTPSSPCGLYHQYHLAGHYASATLWVAFCFWLLSNVLLST
PAPLYGGLALLTTGAFALFGVFALASISSVPLCPLRLGSSALTTQYGAAF
WVTLATGEDRENGPRGLRVETGFTPGVLCLFLGGAVAGKQCPPGLGQESS
RKGTERCWREASDIRRHQGKSPGAICK (SEQ ID NO.: 63)

This protein was analyzed using the SMART program. This analysis indicates that the protein contains three transmembrane domains and a signal sequence. The predicted structure of the protein is depicted schematically below:



WHAT IS CLAIMED IS:

1. An isolated nucleic acid sequence that is expressed by human colon cancer cells selected from the group consisting of:

(i) the nucleic acid sequence contained in SEQ ID NO: 1; 2, 4, 6, 8, 9, 10. 11, 13 and 15

(ii) variants thereof, wherein such variants have a nucleic acid sequence that is at least 70% identical to the sequence of (i) or (ii) when aligned without allowing for gaps; and

(iii) fragments of (i) or (ii) having a size of at least 20 nucleotides in length.

2. The nucleic acid sequence of Claim 1 which comprises the nucleic acid sequence contained in any one of SEQ ID NO: 2, 4, 6, 8, 9, 10, 11, 13 and 15 or a fragment thereof.

3. A primer mixture that comprises primers that result in the specific amplification of one or the cancer genes identified in Claim 1.

4. A method of detecting colon cancer comprising (i) obtaining a human colon cell sample; and (ii) determining whether such cell sample expresses a colon cancer gene having a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 13, 15, 17, 18, 19, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31.

5. The method of Claim 6, wherein said method comprises detecting the expression of said colon cancer gene using a nucleic acid sequence that specifically hybridizes thereto.

6. The method of Claim 5, wherein said method comprises detecting the expression of said colon cancer gene using primers that result in the amplification thereof.

7. The method of Claim 5, wherein the expression of said colon cancer gene is detected by assaying for the antigen encoded by said gene.

8. The method of Claim 7, wherein said assay involves the use of a monoclonal antibody or fragment that specifically binds to said antigen.

9. The method of Claim 8, wherein said assay comprises an ELISA or competitive binding assay.

10. An antigen expressed by human colon cancer cells that is selected from the group consisting of:

(i) the antigen encoded by the nucleic acid sequence in SEQ ID NO. 2, 4, 6, 8, 9, 10, 11, 13 and 15;

(ii) the antigen having the amino acid sequence contained in SEQ ID NO. 5, 7, 11, and 16; and

(iii) fragments or variants thereof that bind to or elicit antibodies that specifically bind the antigen of (i) or (ii).

11. An colon antigen having the amino acid sequence in selected from the group consisting of or an antigen fragment thereof.

12. A monoclonal antibody or antigen-binding fragment thereof that specifically binds to an antigen according to Claim 10 or 11.

13. A monoclonal antibody or fragment that specifically binds the antigen of Claim 12.

14. The antigen of Claim 10 or 11 which is attached directly or indirectly to a detectable label.

15. The antibody of Claim 12 or 13 which is attached directly or indirectly to a detectable label.

16. A diagnostic kit for detection of colon cancer which comprises a DNA according to Claim 1 and a detectable label.

17. A diagnostic kit for detection of colon cancer which comprises primers according to Claim 3 and a diagnostically acceptable carrier.

18. A diagnostic kit for detection of colon cancer which comprises a monoclonal antibody according to Claim 12 or 13 and a detectable label.

19. A method for treating colon cancer which comprises administering a therapeutically effective amount of a ribozyme or antisense oligonucleotide that inhibits the expression of a gene having a DNA sequence selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 9, 10, 11, 13, 15, 17, 18, 19, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31 or a fragment, or variant thereof.

20. A method for treating colon cancer which comprises administering a nucleic acid sequence that specifically binds a gene selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 9, 10, 11, 13, 15, 17, 18, 19, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31 or a fragment, or variant thereof which is directly or indirectly attached to an effector moiety.

21. The method of Claim 20, wherein said effector moiety is a therapeutic radiolabel, enzyme, cytotoxin, growth factor, or drug.

22. A method for treating colon cancer comprising administering a therapeutically effective amount of an antigen according to Claim 12 or 13 and an adjuvant that elicits a humoral or cytotoxic T-lymphocyte response to said antigen.

23. A method for treating colon cancer comprising administering a therapeutically effective amount of a ligand which specifically binds to a protein encoded by gene having a sequence selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 9, 10, 11, 13, 15, 17, 18, 19, 20, 22, 23, 24, 26, 27, 28, 29, 30, 31 or a fragment, or variant thereof optionally directly or indirectly attached to a therapeutic effector moiety.

24. The method of Claim 23, wherein said effector moiety is a radiolabel, enzyme, cytotoxin, growth factor, or drug.

25. The method of Claim 24 wherein the radiolabel is yttrium.

26. The method of Claim 25 wherein the radiolabel is indium.

27. The method of claim 23 wherein said ligand is a monoclonal antibody or fragment thereof.

28. The method of claim 23 wherein said ligand is a small molecule.

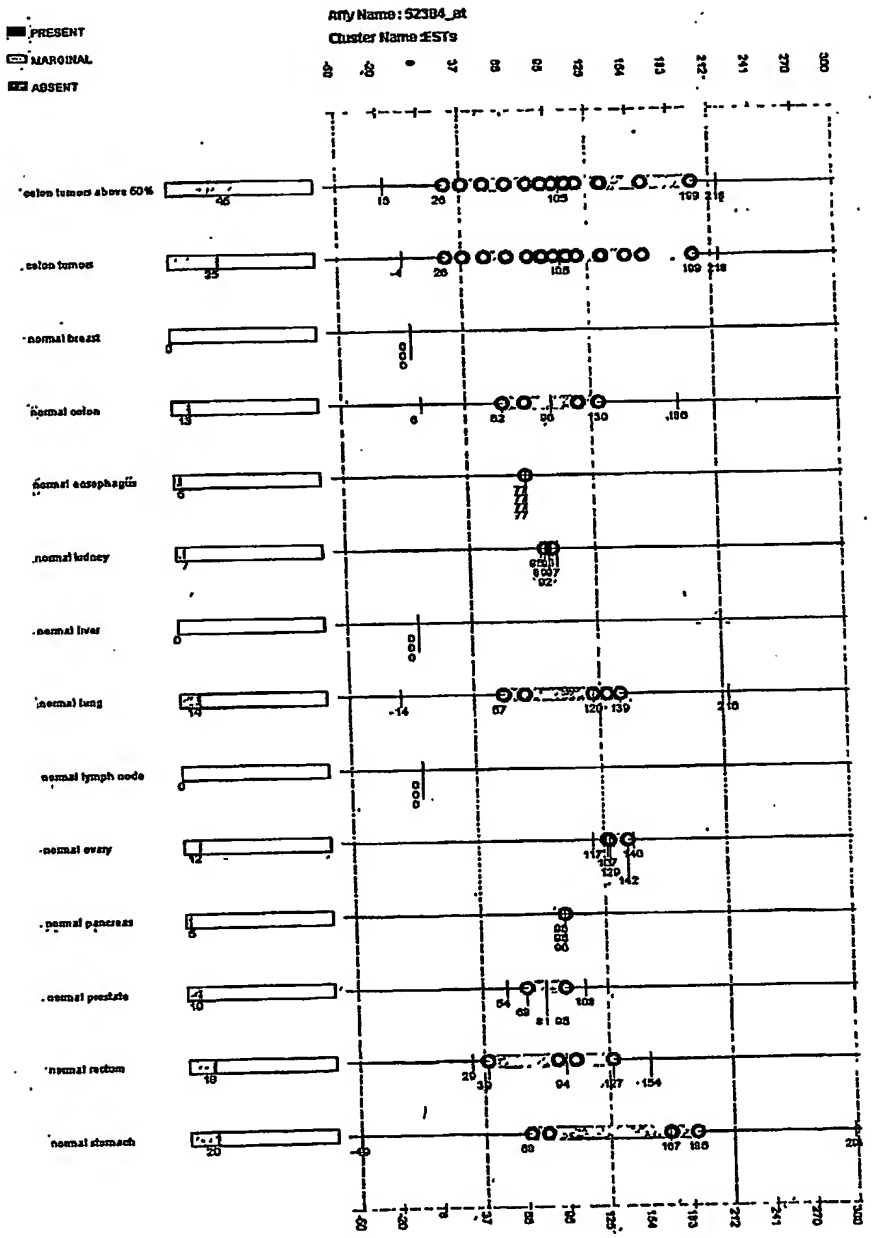
29. The method of claim 23 wherein said ligand is a peptide.

ABSTRACT OF THE DISCLOSURE

The invention identifies a number of genes that are overexpressed in colon or colorectal tumor tissues. These genes and the corresponding antigens are useful diagnostic and therapeutic targets.

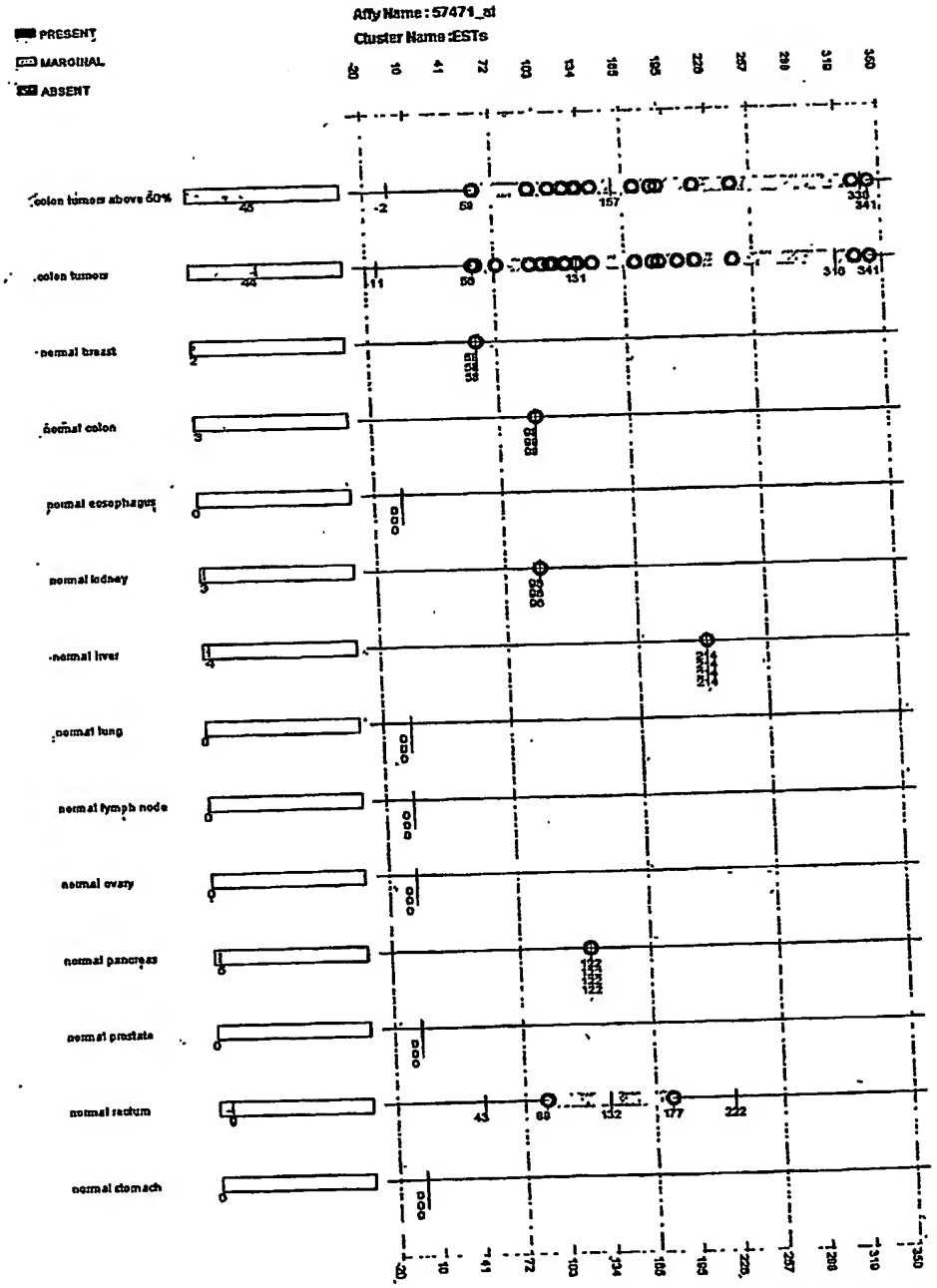
Candidate 1: W91975

FIGURE 1



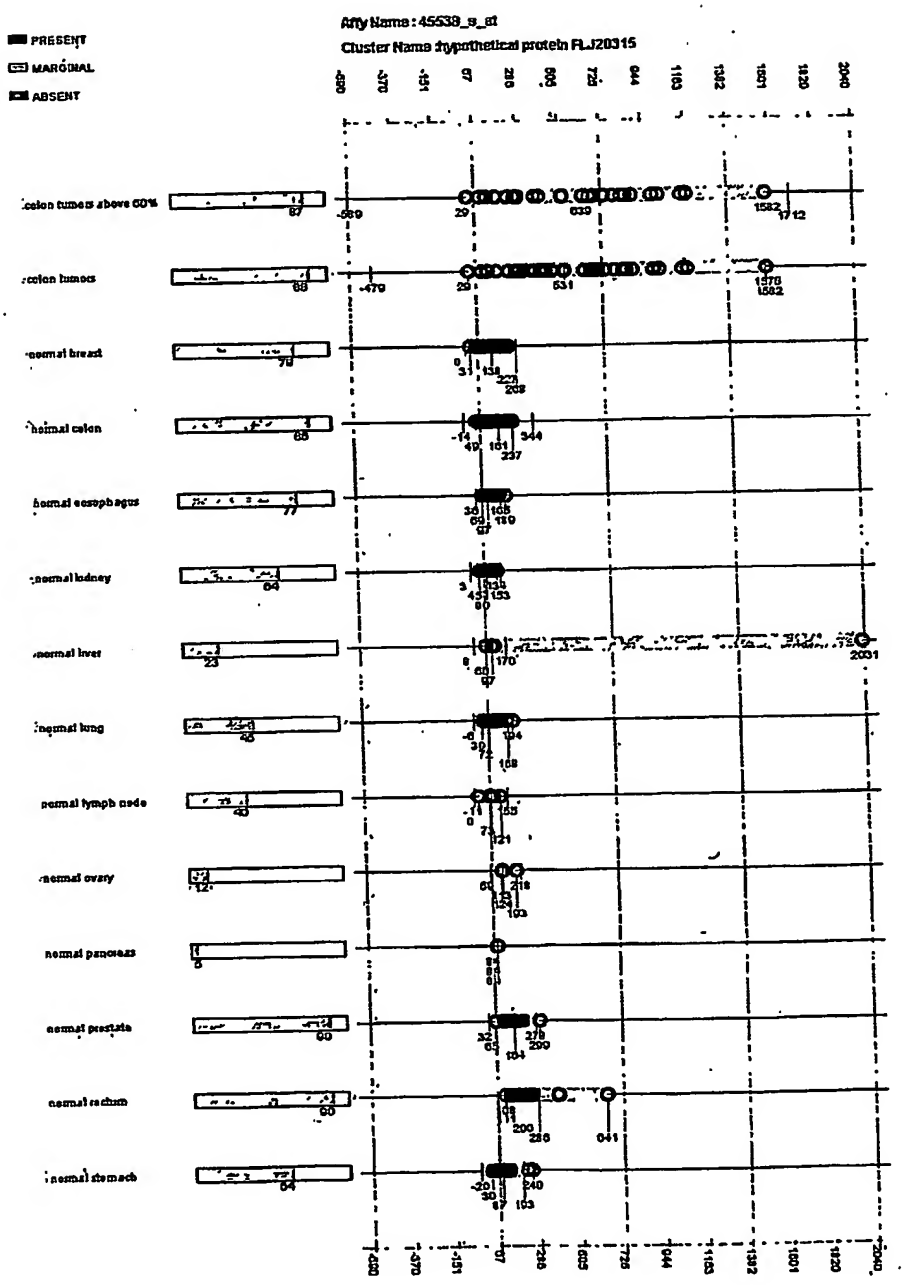
Candidate 2: AI694242

FIGURE 2



Candidate 3: AI1680111

FIGURE 3



Candidate 4: AA813827

FIGURE 4

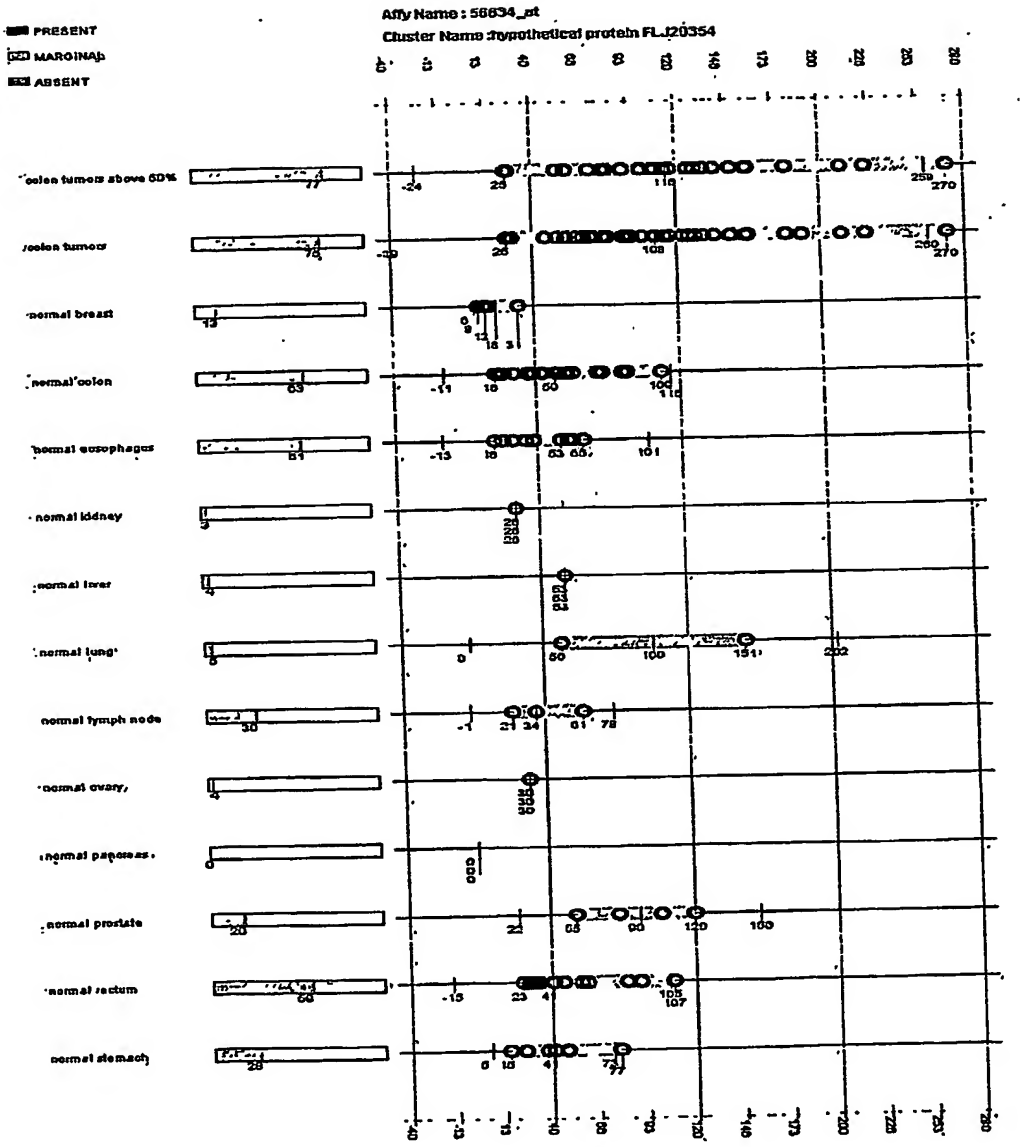


FIGURE 5

Expression of Malignant colon specific Sequences

Name	Designation	Size	A Tumor	A Tumor	AN:N	A Normal	BT:N	B Tumor	Pool	Pooled Normal	DT:N	D Tumor
CICO1	bs210ms130	184.75	3	1213.04	3	256.05	3	2685.32	2	298.09	3	1812.48
CICO2	bs220ms230	191.20	3	994.12	3	108.37	3	228.49	3	60.10	3	1271.98
CICO3	bs430ms430	222.32	3	1587.80	3	113.79	3	1226.02	3	117 06	3	163.29
	bs421ms433	258.00	3	213 22	3	62.02	3	175.93	2	36.16	3	158.55

FT2:N	F Tumor	FN2:N	F Normal	GT:N	G Tumor	GN:N	G Normal	HT:N	H Tumor	HN:N	H Normal	IT:N	I Tumor
3	658.97	3	229.01	3	2750.16	3	307.80	3	858.10	3	271.62	3	2501.17
3	370.06	3	41.80	3	590.73	3	56.69	3	295.42	3	90.93	3	1164.92
3	2747.40	3	118.80	3	1559.78	3	112.18	3	302.67	3	66.51	3	8010.81
3	148 52	3	45 92	3	306 88	3	115.65	3	168 77	3	75.19	3	190 42

IN:N	I Normal	MT:N	M Tumor	MN:N	M Normal	QT:N	Q Tumor	QN:N	Q Normal	RT:N	R Tumor	RN:N	R Normal
3	225.34	3	3354.80	3	267.40	3	1494.28	3	261.62	3	2518.73	3	550.77
3	57.46	3	90.37	3	47.53	3	132.73	2	66.31	3	376.15	3	55.26
3	90.56	2	280.45	3	245.67	3	2593.52	3	158.56	3	1061 67	3	70 30
3	29 94	3	95 06	2	34.97	3	237 60	3	24.70	3	293 87	3	53 45

8T:N	8 Tumor	Tumor Ave	Normal Ave	Ave Fold	# of hits
3	256.60	1827.61	296.41	6.17	20
1	130.50	513.22	64.94	7.90	35
3	160.96	1790.40	121.49	14.74	19
3	248.80	203.42	53.11	3.83	22

SCANNED #5

Figure 6

E Northern Data

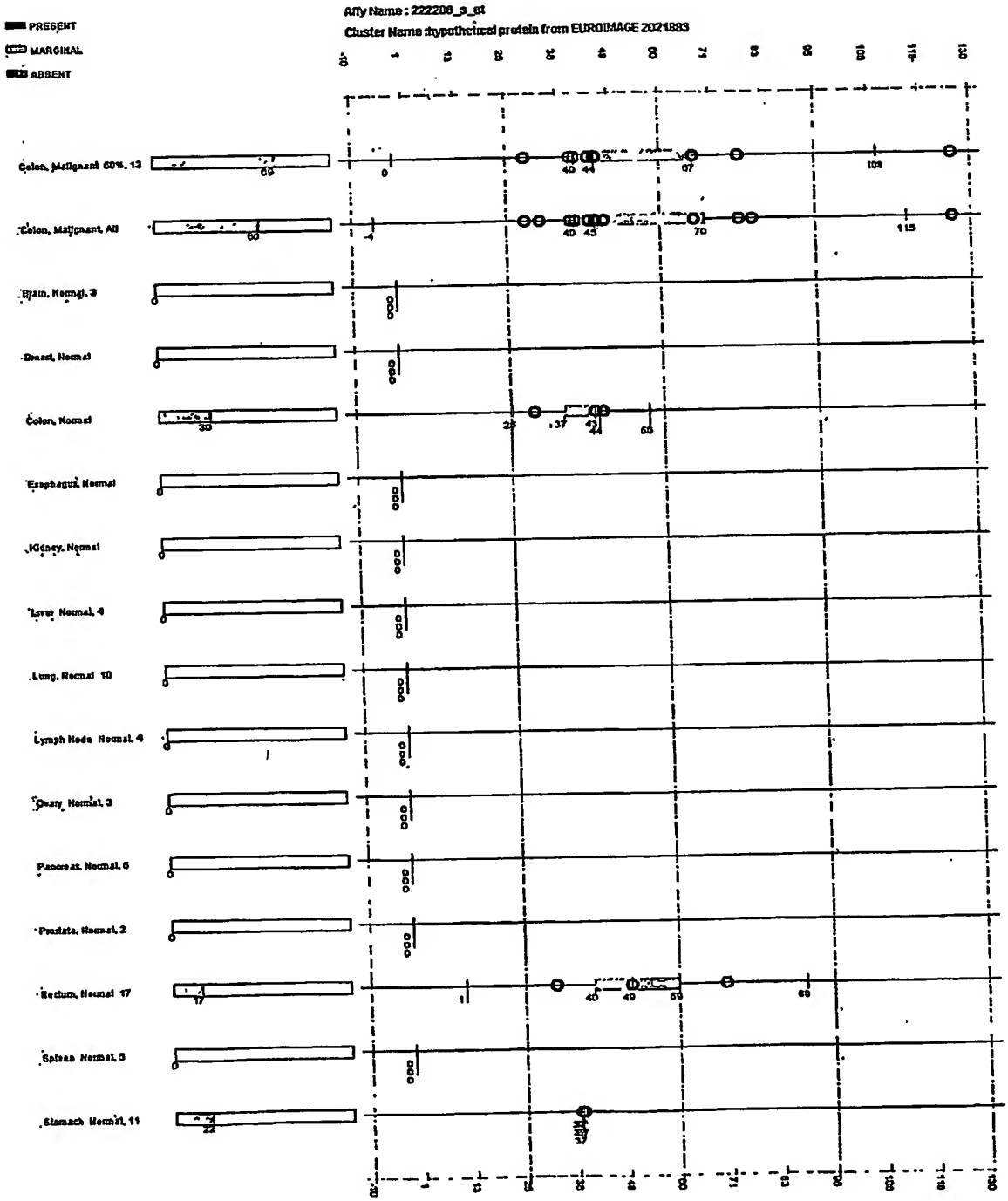


Figure 7

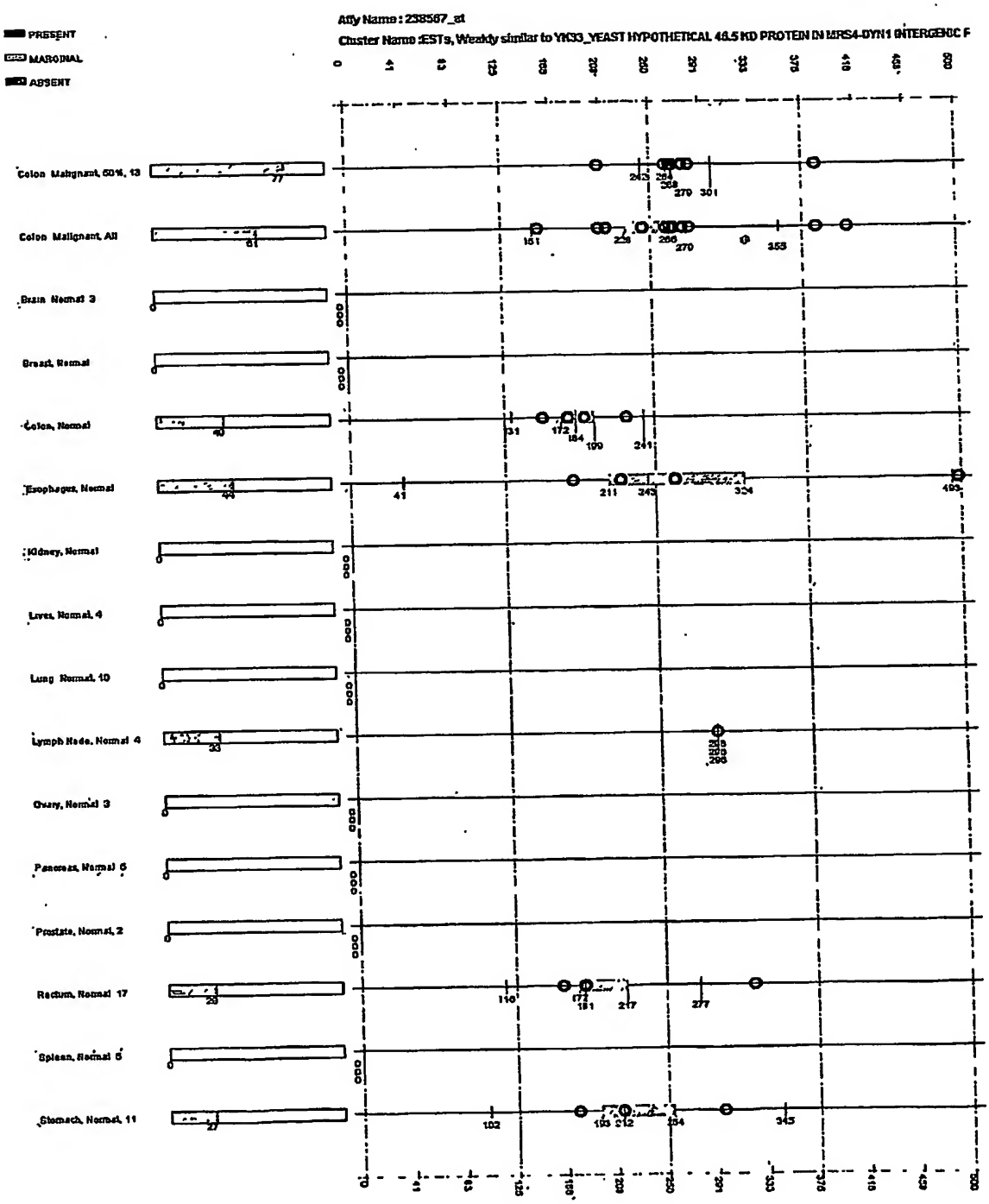

Affy Name : 238567_at
Cluster Name :ESTs, Weakly similar to YH33_YEAST HYPOTHETICAL 46.5 KD PROTEIN IN MRS4-DYN1 INTERGENIC F
PRESENT
MARGINAL
ABSENT
Colon Malignant, 50%, 13
Colon Malignant, All
Brain Normal 3
Breast, Normal
Colon, Normal
Esophagus, Normal
Kidney, Normal
Liver, Normal, 4
Lung Normal, 10
Lymph Node, Normal 4
Ovary, Normal 3
Pancreas, Normal 5
Prostate, Normal, 2
Rectum, Normal 17
Spleen, Normal 5
Stomach, Normal, 11

Figure 8

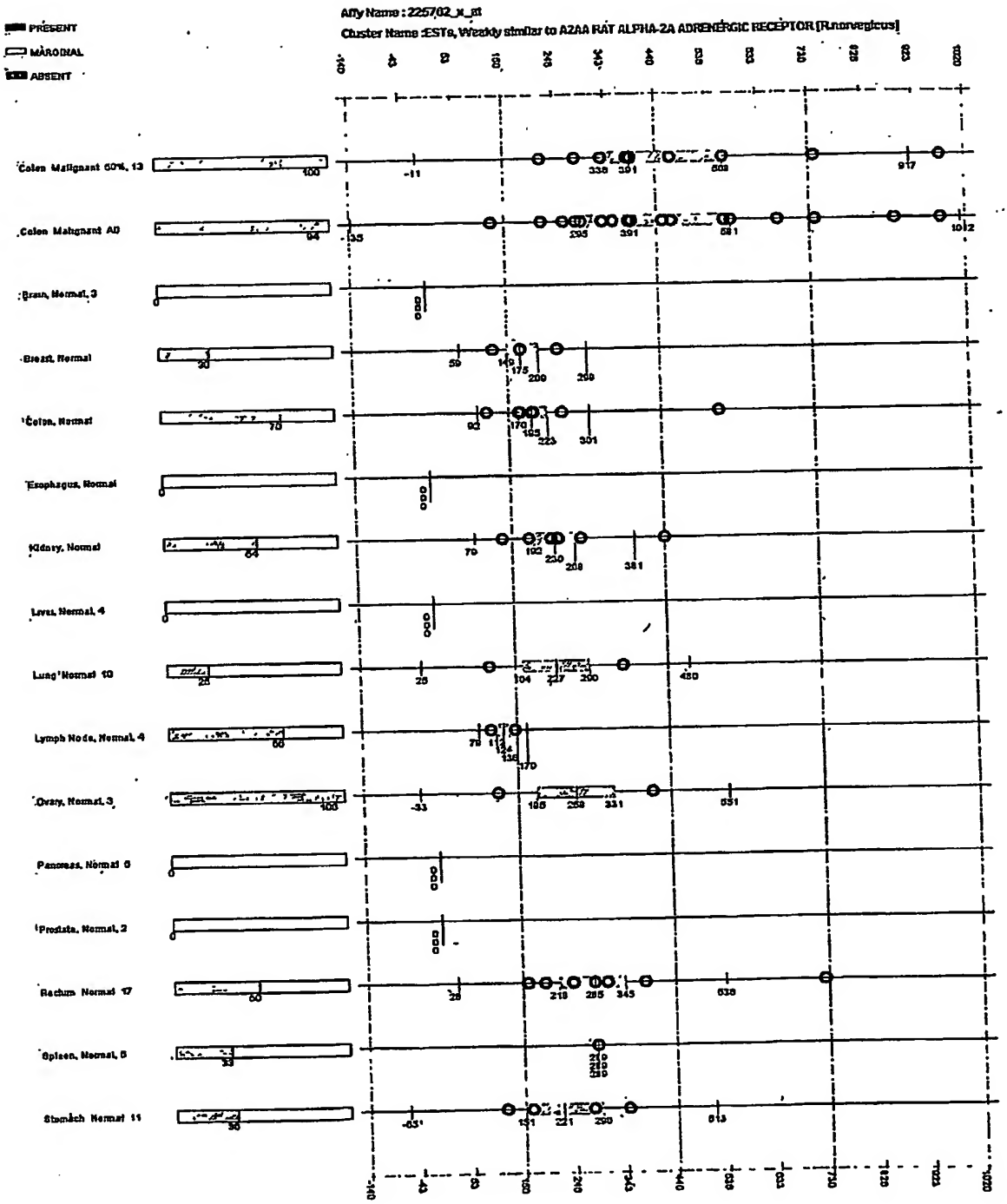

PRESENT
MARGINAL
ABSENT
Affy Name : 225702_x_at
Cluster Name :ESTs, Weakly similar to A2AA RAT ALPHA-2A ADRENERGIC RECEPTOR [Rnorvegicus]
Colon Malignant 60%, 13
Colon Malignant A0
Brain, Normal, 3
Breast, Normal
Colon, Normal
Esophagus, Normal
Kidney, Normal
Liver, Normal, 4
Lung Normal 10
Lymph Node, Normal, 4
Ovary, Normal, 3
Pancreas, Normal 0
Prostate, Normal, 2
Rectum Normal 17
Spleen, Normal, 0
Stomach Normal 11

Figure 9

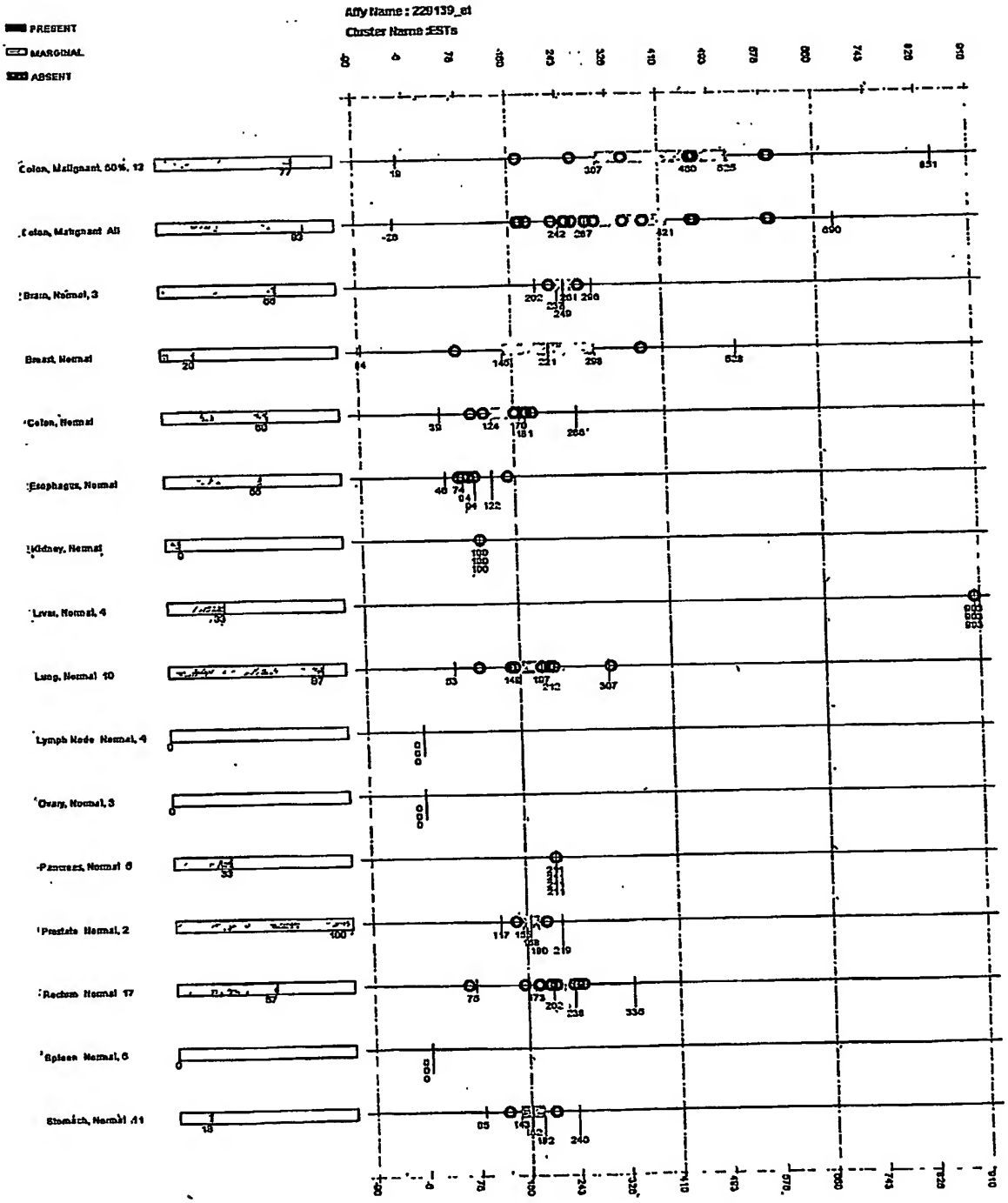

Affy Name : 229139_at
Cluster Name :ESTs
PRESENT
MARGINAL
ABSENT
Colon, Malignant, 50%, 12
Colon, Malignant All
Brain, Normal, 3
Breast, Normal
Colon, Normal
Esophagus, Normal
Kidney, Normal
Liver, Normal, 4
Lung, Normal 10
Lymph Node Normal, 4
Ovary, Normal, 3
Pancreas, Normal 6
Prostate Normal, 2
Rectum Normal 17
Spleen Normal, 6
Stomach, Normal 11

Figure 10

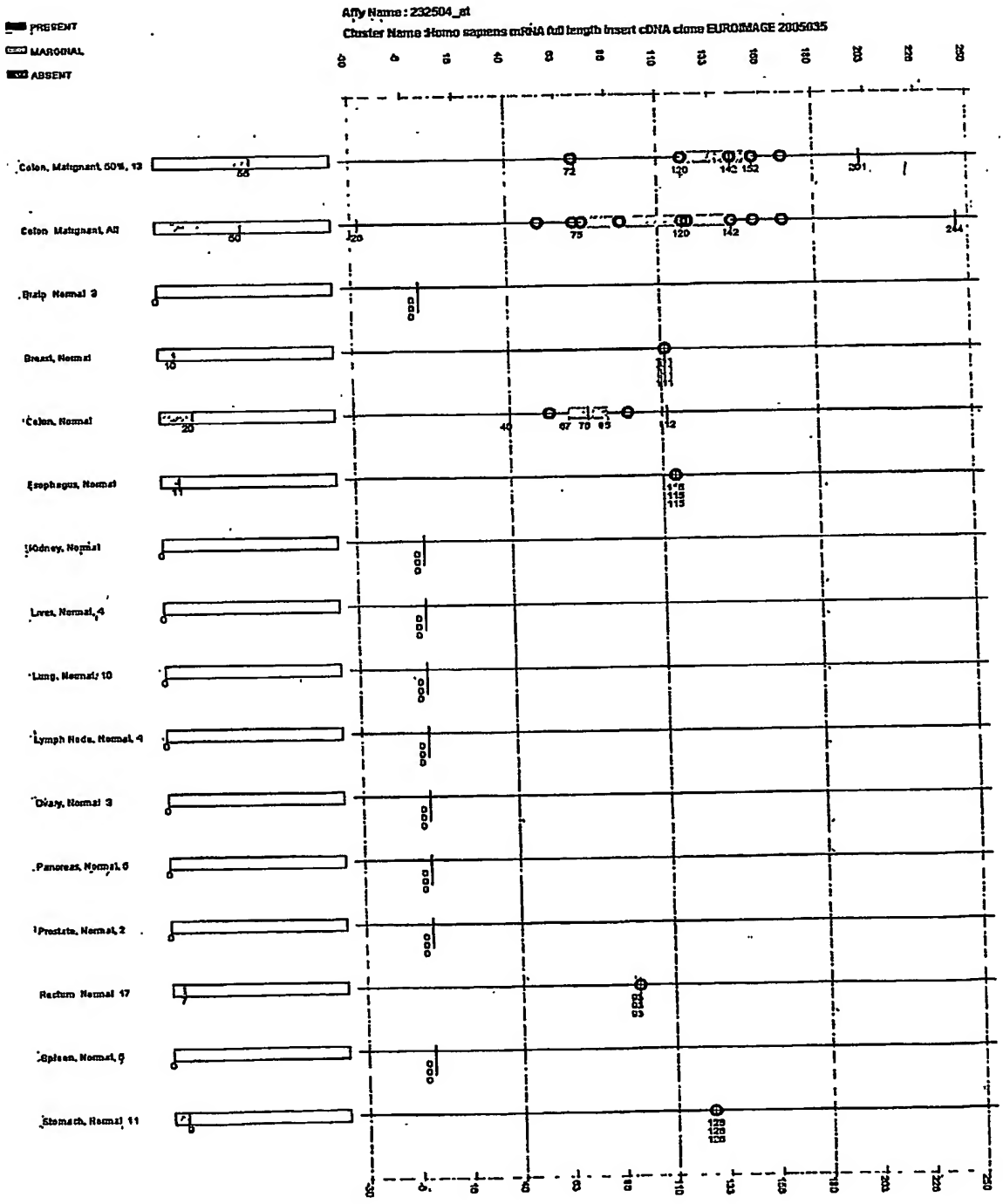
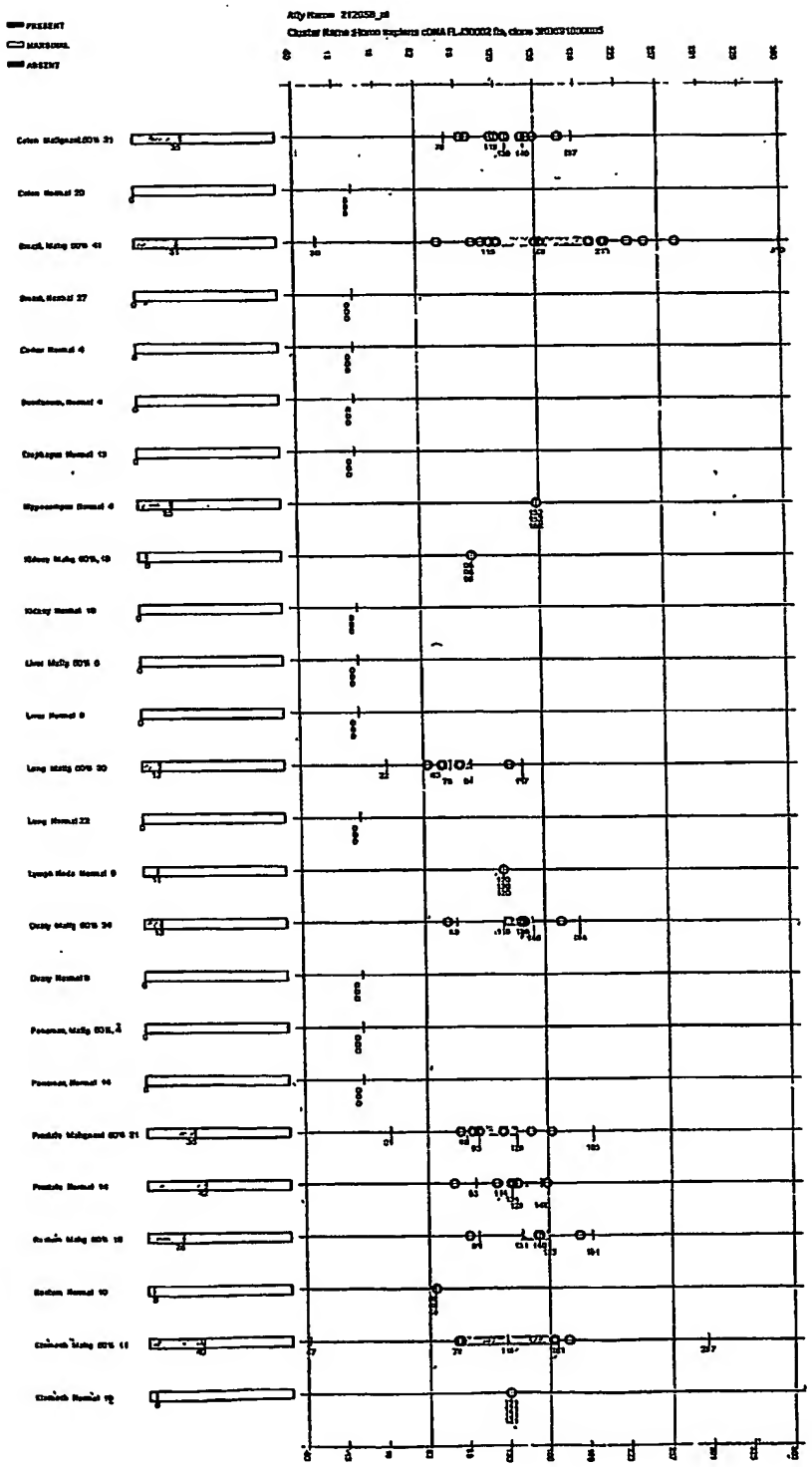

PRESENT
MARGINAL
ABSENT
Affy Name : 232504_at
Cluster Name :Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 2005035
Colon, Malignant, 60%, 13
Colon Malignant, All
Brain Normal 3
Breast, Normal
Colon, Normal
Esophagus, Normal
Kidney, Normal
Liver, Normal, 4
Lung, Normal, 10
Lymph Node, Normal, 4
Ovary, Normal 3
Pancreas, Normal, 6
Prostate, Normal, 2
Rectum Normal 17
Spleen, Normal, 6
Stomach, Normal 11

Figure 11



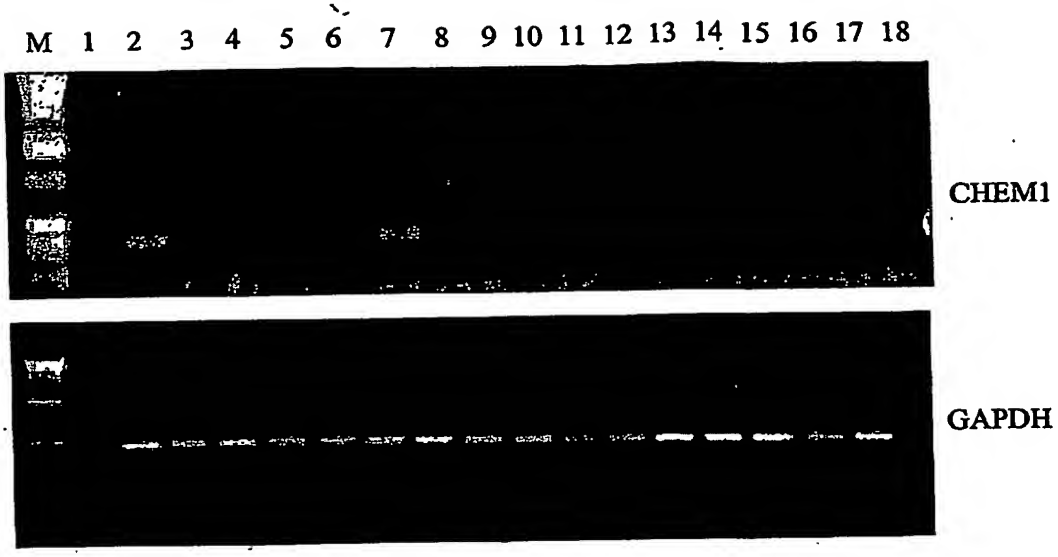


Figure 12 CHEM1 message in multi-tissue panel 1. 1 ng of cDNA from 1 no cDNA, 2 prostate tumor N, 3 prostate tumor O, 4 prostate tumor T, 5 colon tumor f, 6 colon tumor G, 7 colon tumor R, 8 normal brain, 9 normal colon, 10 normal heart, 11 normal kidney, 12 normal liver, 13 normal lung, 14 normal skeletal muscle, 15 normal pancreas, 16 normal placenta, 17 normal prostate, 18 normal thymus.

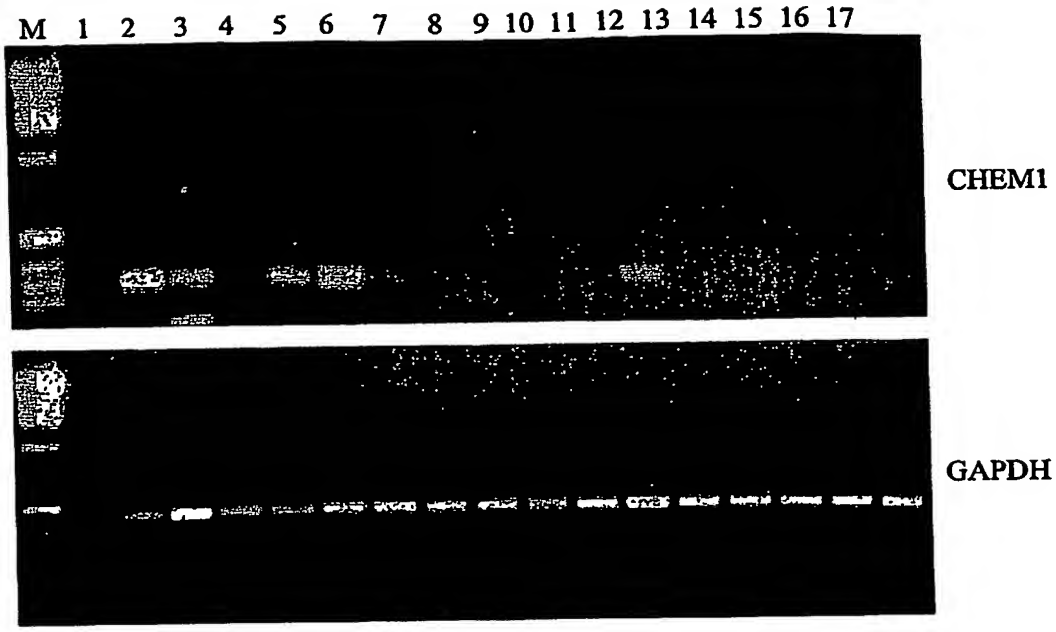


Figure 13 CHEM1 message in multi-tissue panel 1. 5 ng of cDNA from 1 no cDNA, 2 prostate tumor N, 3 prostate tumor O, 4, colon tumor f, 5 colon tumor G, 6 colon tumor R, 7 normal brain, 8 normal colon, 9 normal heart, 10 normal kidney, 11 normal liver, 12 normal lung, 13 normal skeletal muscle, 14 normal pancreas, 15 normal placenta, 16 normal prostate, 17 normal thymus.

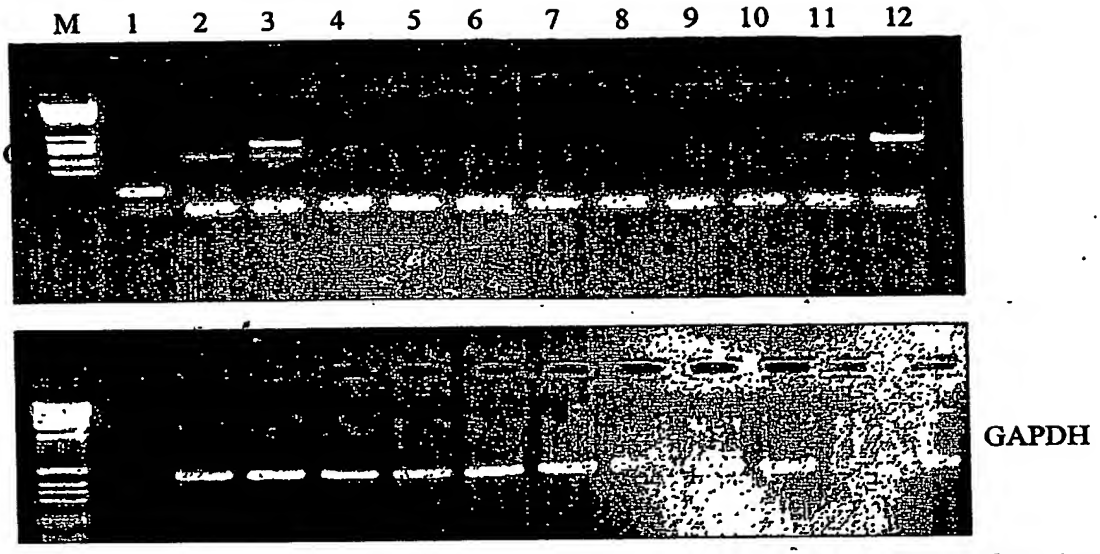


Figure 14 CHEM1 message in multi-tissue panel II. 5 ng of cDNA from 1 no cDNA, 2 prostate tumor N, 3 colon tumor R, 4, normal colon, 5 normal heart, 6 normal peripheral blood lymphocytes, 7 normal small intestine, 8 normal ovary, 9 normal spleen, 10 normal testis, 11 normal thymus.

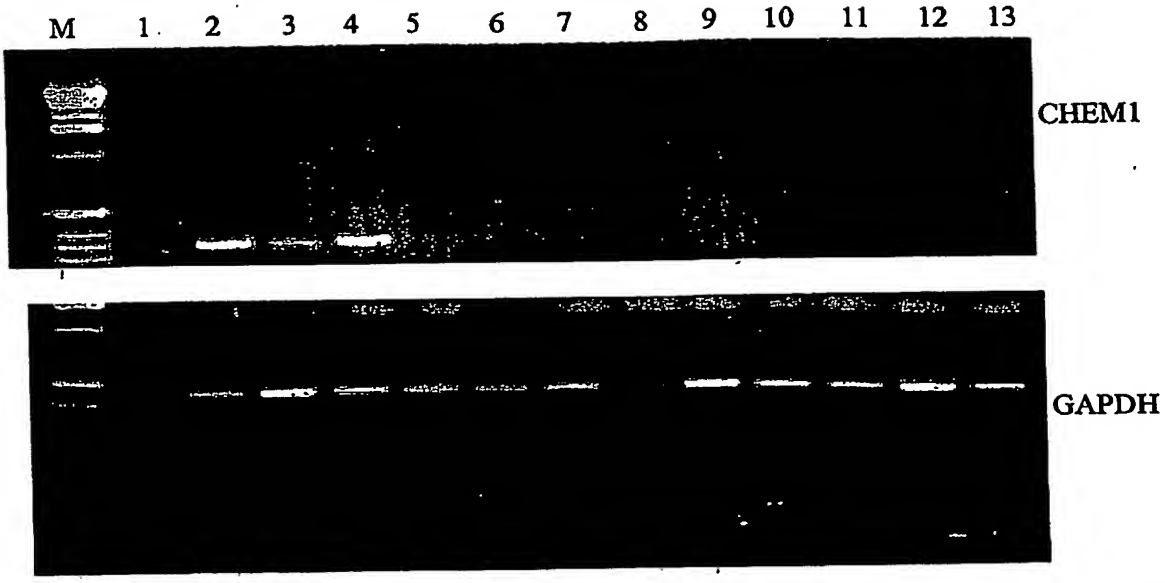


Figure 15 CHEM1 message in brain tissue panel. 5 ng of cDNA from 1 no cDNA, 2 prostate tumor N, 3 prostate tumor O, 4, colon tumor R, 5 cerebral cortex, 6 cerebellum, 7 medulla oblongata, 8 pons, 9 frontal lobe, 10 occipital lobe, 11 parietal lobe, 12 temporal lobe, 13 placenta.

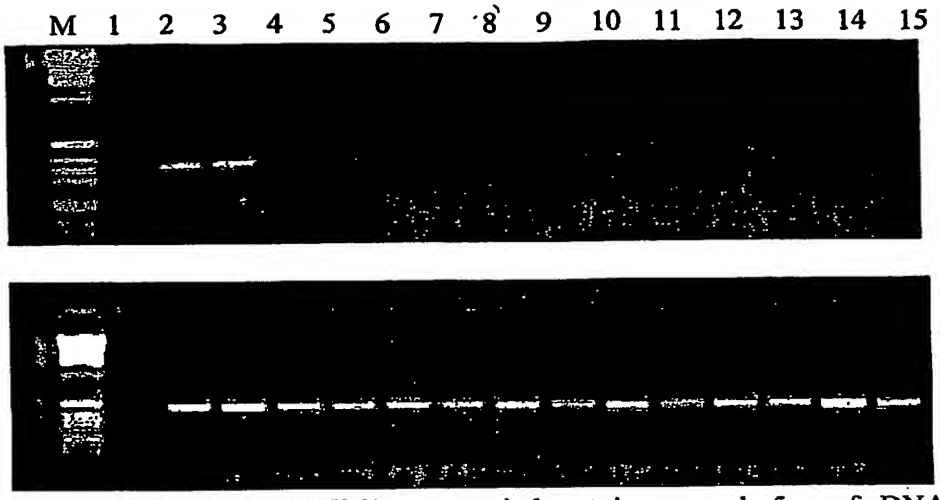
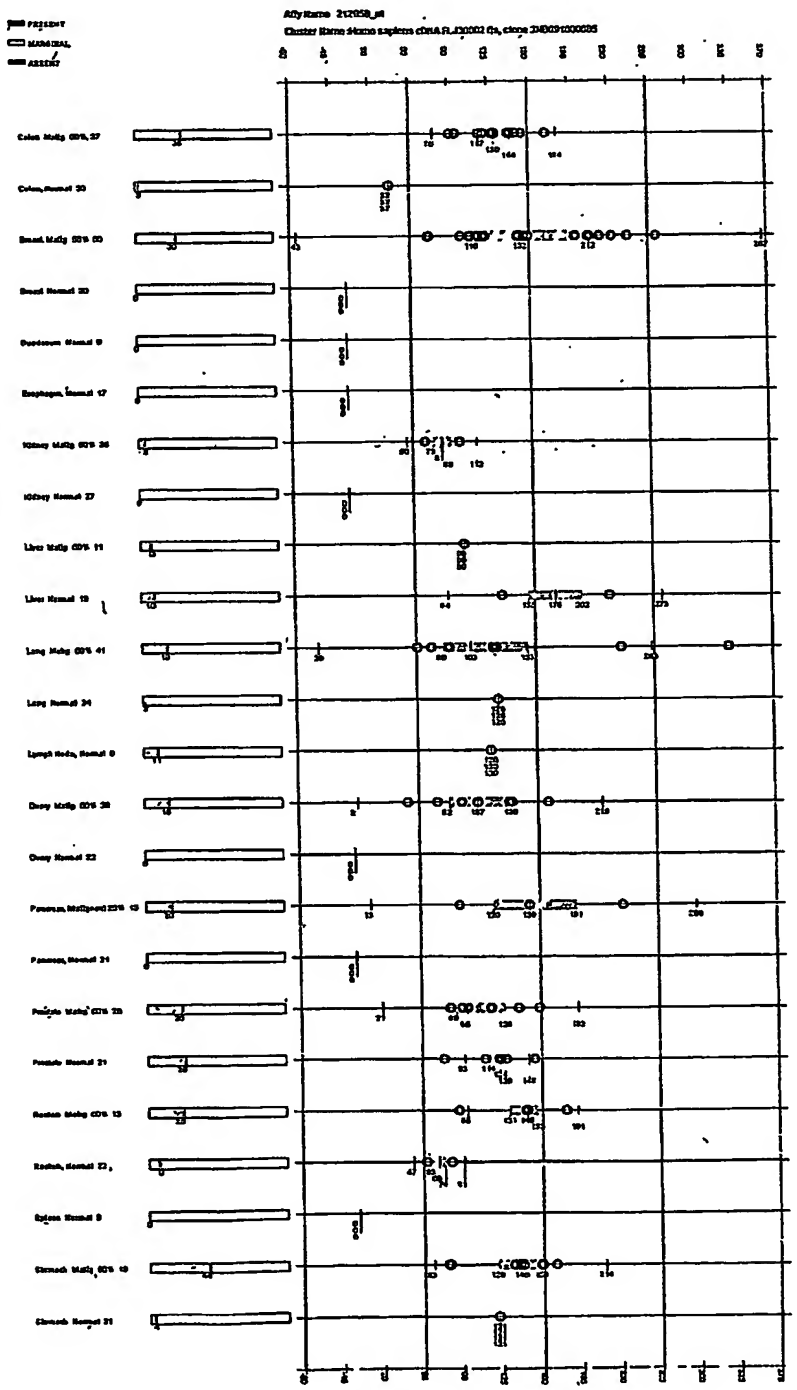


Figure 16 CHEM1 message in heart tissue panel. 5 ng of cDNA from 1 no cDNA, 2 prostate tumor N, 3 colon tumor R, 4 adult heart, 5 fetal heart, 6 aorta, 7 apex, 8 left atrium, 9 right atrium, 10 left ventricle, 11 right ventricle, 12 dextra auricle, 13 sinistra auricle, 14 atrioventricular node, 15 septum intraven.

Figure 17 contains additional Enorthern results showing that this protein is expressed in lung, ovarian, pancreatic, breast, colon, stomach and prostate cancers.



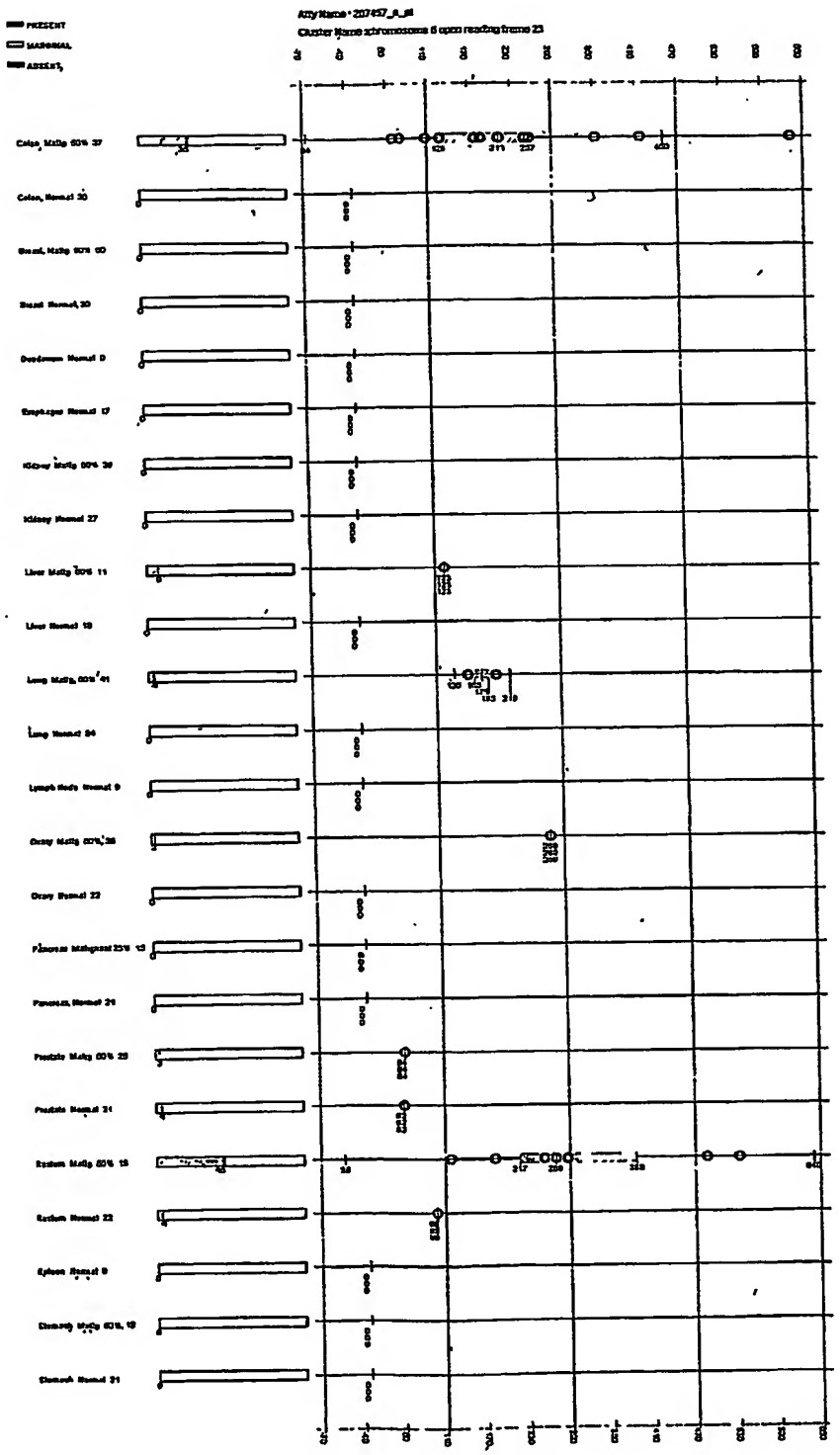


Figure 18

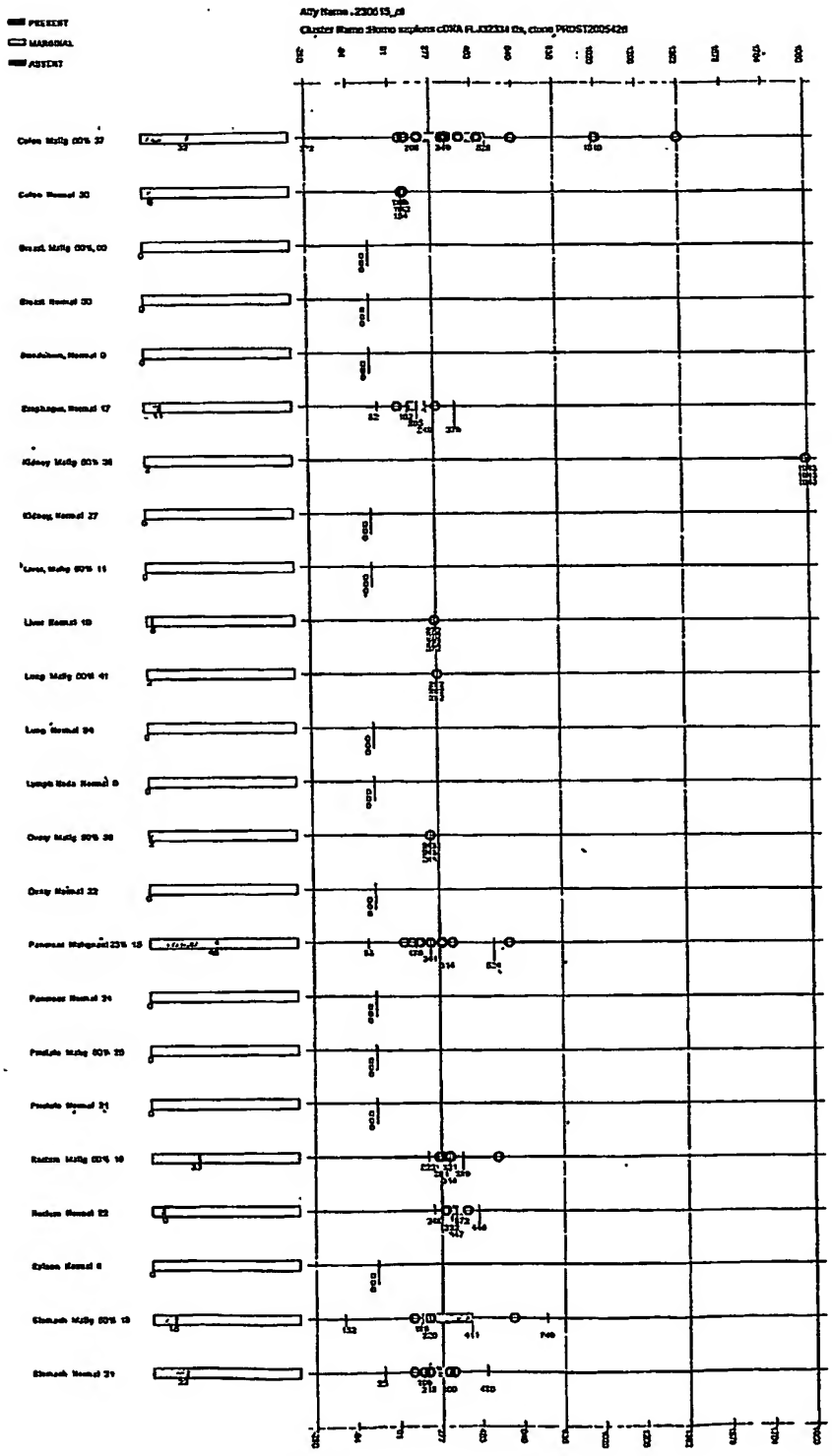


Figure 19